SUPPORT FOR THE AMENDMENTS

The specification has been amended to insert the address of the depository and to replace the Abstract. The amendments to the claims are supported by the specification. Accordingly, no new matter is believed to have been added to the present application by the amendments submitted above.

REMARKS

Claims 86-104 and 106-118 are pending. Favorable reconsideration is respectfully requested.

The rejections of the claims under 35 U.S.C. §112, first paragraph, are respectfully traversed.

The present specification provides a detailed description of the procedure for conducting the claimed method. In view of that description, one would appreciate that the invention as claimed is described and could be practiced with routine experimentation.

In addition, Applicants submit herewith publications from the scientific literature which demonstrate that valyl-tRNA synthetase genes had been extensively reported prior to the filing date of the present application.

Jordana et al. (*J. Biol. Chem.* (1987) 262(15): 7189-94) demonstrate that as early as 1987 the sequence of the valyl-tRNA synthetase gene in *Saccharomyces cerevisiae* (yeast) and reported a high level of homology between both yeast and bacteria aminoacyl-RNA genes.

Heck et al. (*J. Biol. Chem.* (1988) 26(2): 868-877) disclosed in 1988 the cloning and sequencing of ValS of E. coli. These authors also found that ValS was highly related with yeast valyl-tRNA genes.

Brown et al. (PNAS (1995) 92:2441-45) sequenced, in 1995, isoleucyl-tRNA genes in a large array of microbial species and demonstrated that this family of genes probably expanded through the species by duplication.

Luo et al. (J. Bact. (1997) 179(8): 2472-2478) sequenced ValS in *B. subtilis* and found further similarities between valyl-tRNA synthetases from *Bacillus subtilis*, *Bacillus stearothermophilus*, *Lactobacillus casei* and *Escherichia coli*.

The publications submitted herewith demonstrate that it was well-established in the art at the time the present application was filed that valyl-tRNA synthetase genes shared strong similarities throughout bacteria and yeast.

In view of the foregoing, the present specification describes and enables the claimed method. Accordingly, withdrawal of this ground of this ground of rejection is respectfully requested.

Regarding the issue with respect to biological deposit of the subject matter of Claim 106, Applicants confirm that the deposits were made under the terms of the Budapest Treaty. Copies of the deposit receipts are submitted herewith. The complete address for the depository has been added to the specification.

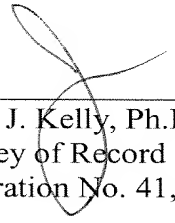
Withdrawal of this ground of rejection is respectfully requested.

The rejection of the claims under 35 U.S.C. §112, second paragraph, is believed to be obviated by the amendment submitted above. The claims have been amended as suggested by the Examiner in order to address the issues raised in the Office Action. In view of the foregoing, the claims are definite within the meaning of 35 U.S.C. §112, second paragraph. Accordingly, withdrawal of this ground of rejection is respectfully requested.

Applicants submit that the present application is in condition for allowance. Early notice to this effect is earnestly solicited.

Respectfully submitted,

OBLON, SPIVAK, McCLELLAND,
MAIER & NEUSTADT, P.C.



James J. Kelly, Ph.D.
Attorney of Record
Registration No. 41,504

Customer Number
22850

Tel: (703) 413-3000
Fax: (703) 413 -2220
(OSMMN 06/04)

TRAITE DE BUDAPEST SUR LA RECONNAISSANCE
INTERNATIONALE DU DEPOT DES MICRO-ORGANISMES
AUX FINS DE LA PROCEDURE EN MATIERE DE BREVETS

FORMULE INTERNATIONALE

DESTINATAIRE :

INSTITUT PASTEUR
Bureau des Brevets et Inventions
25-28, rue du Docteur Roux
75015 PARIS

RECEPISSE EN CAS DE DEPOT INITIAL,
délivré en vertu de la règle 7.1 par
l'AUTORITE DE DEPOT INTERNATIONALE
identifiée au bas de cette page

NOM ET ADRESSE
DU DEPOSANT

I. IDENTIFICATION DU MICRO-ORGANISME

Référence d'identification donnée par le DEPOSANT :	Numéro d'ordre attribué par l'AUTORITE DE DEPOT INTERNATIONALE :
β5485	**I - 2340**

II. DESCRIPTION SCIENTIFIQUE ET/OU DESIGNATION TAXONOMIQUE PROPOSEE

Le micro-organisme identifié sous chiffre I était accompagné :

■ d'une description scientifique

■ d'une désignation taxonomique proposée

(Cocher ce qui convient)

III. RECEPTION ET ACCEPTATION

La présente autorité de dépôt internationale accepte le micro-organisme identifié sous chiffre I, qu'elle a reçu le **26 OCTOBRE 1999** (date du dépôt initial) [1]

IV. RECEPTION D'UNE REQUETE EN CONVERSION

La présente autorité de dépôt internationale a reçu le micro-organisme identifié sous chiffre I le (date du dépôt initial)
et a reçu une requête en conversion du dépôt initial en dépôt conforme au Traité de Budapest le (date de réception de la requête en conversion)

V. AUTORITE DE DEPOT INTERNATIONALE

Nom : **CNCM**
Collection Nationale
de Cultures de Microorganismes

Adresse : **INSTITUT PASTEUR**
28, Rue du Docteur Roux
F-75724 PARIS CEDEX 15

Signature(s) de la (des) personne(s) compétente(s) pour représenter l'autorité de dépôt internationale ou de l'(des) employé(s) autorisé(s) : **Simona OZDEN**
Directeur de la CNCM

Date : **Paris, le 30 novembre 1999**

1 En cas d'application de la règle 6.4.d), cette date est la date à laquelle le statut d'autorité de dépôt internationale a été acquis.

Formule BP/4 (page unique)

TRAITE DE BUDAPEST SUR LA RECONNAISSANCE
INTERNATIONALE DU DEPOT DES MICRO-ORGANISMES
AUX FINS DE LA PROCEDURE EN MATIERE DE BREVETS

FORMULE INTERNATIONALE

DESTINATAIRE :

Madame Danielle BERNEMAN,
Bureau des Brevets et Inventions
INSTITUT PASTEUR
25-28, rue du Docteur Roux
75724 PARIS CEDEX 15

NOM ET ADRESSE DE LA PARTIE A LAQUELLE LA DECLARATION SUR LA VIABILITE EST DELIVREE

DECLARATION SUR LA VIABILITE,
délivrée en vertu de la règle 10.2 par
l'AUTORITE DE DEPOT INTERNATIONALE
identifiée à la page suivante

I. DEPOSANT	II. IDENTIFICATION DU MICRO-ORGANISME
Nom : **INSTITUT PASTEUR** Adresse : **Bureau des Brevets et Inventions** **25-28, rue du Docteur Roux** **75015 PARIS**	Numéro d'ordre attribué par l'AUTORITE DE DEPOT INTERNATIONALE : **I - 2340** Date du dépôt ou du transfert [1] : **26 OCTOBRE 1999**

III. DECLARATION SUR LA VIABILITE

La viabilité du micro-organisme identifié sous chiffre II a été contrôlée le **27 OCTOBRE 1999** [2]. A cette date, le micro-organisme

[■] [3] était viable

[ ] [3] n'était plus viable

1 Indiquer la date du dépôt initial ou, si un nouveau dépôt ou un transfert ont été effectués, la plus récente des dates pertinentes (date du nouveau dépôt ou date du transfert).

2 Dans les cas visés à la règle 10.2.a)ii) et iii), mentionner le contrôle de viabilité le plus récent.

3 Cocher la case qui convient.

IV. CONDITIONS DANS LESQUELLES LE CONTROLE DE VIABILITE A ETE EFFECTUE [4]

V. AUTORITE DE DEPOT INTERNATIONALE

Nom :

CNCM
Collection Nationale
de Cultures de Microorganismes

Adresse :

INSTITUT PASTEUR
28, Rue du Docteur Roux
F-75724 PARIS CEDEX 15
FRANCE

Signature(s) de la (des) personne(s) compétente(s) pour représenter l'autorité de dépôt internationale ou de l'(des) employé(s) autorisé(s) :

Simona OZDEN
Directeur de la CNCM

Georges WAGENER
Conseiller Scientifique de la CNCM
pour les bactéries

Date : Paris, le 30 novembre 1999

4 A remplir si cette information a été demandée et si les résultats du contrôle étaient négatifs.

BUDAPEST TREATY ON THE INTERNATIONAL
RECOGNITION OF THE DEPOSIT OF MICROORANISMS
FOR THE PURPOSES OF PATENT PROCEDURE

INTERNATIONAL SYSTEM

RECIPIENTS:

INSTITUT PASTEUR
Bureau des Brevets et Inventions
25-28, rue du Docteur Roux
75015 PARIS

RECEIPT FOR INITIAL DEPOSIT,
issued in accordance with rule 7.1 by the
INTERNATIONAL DEPOSIT AUTHORITY
identified at the bottom of this page

NAME AND ADDRESS OF
DEPOSITOR

I. IDENTIFICATION OF THE MICROORGANISM	
Identification reference given by the DEPOSITOR **β5485**	Serial number given by the INTERNATIONAL DEPOSIT AUTHORITY **I-2340**

II. SCIENTIFIC DESCRIPTION AND/OR PROPOSED TAXONOMIC DESIGNATION

The microorganism identified under heading I was accompanied:

■ By a scientific description

■ By a proposed taxonomic description

(Check the appropriate box)

III. RECEIPT AND ACCEPTANCE

The present International Deposit Authority accepts the microorganism identified under heading I, which it received on **October 26, 1999** (date of the initial deposit)[1]

IV. RECEIPT OF A REQUEST FOR CONVERSION

The present International Deposit Authority received the microorganism identified under heading I, which it received on (date of the initial deposit)[1]
and received a conversion request of the initial deposit into a deposit which conforms to the Budapest Treaty on (date of the receipt of conversion request)

V. INTERNATIONAL DEPOSIT AUTHORITY

Name : **CNCM** Collection Nationale de Cultures De Microorganismes Address : **INSTITUT PASTEUR** 28, rue du Docteur Roux F-75724 PARIS CEDEX 15	Signature(s) of the person(s) competent to represent the International Deposit Authority or the authorized employee(s): **Simona OZDEN** Director of CNCM [signature] Date : **Paris, November 30, 1999**

1. In the case of application of rule 6.4.d), this date is the date on which the authorizing statute for international deposit was acquired.

BUDAPEST TREATY ON THE INTERNATIONAL
RECOGNITION OF THE DEPOSIT OF MICROORANISMS
FOR THE PURPOSES OF PATENT PROCEDURE

INTERNATIONAL SYSTEM

RECIPIENTS:

INSTITUT PASTEUR
Bureau des Brevets et Inventions
25-28, rue du Docteur Roux
75015 PARIS

DECLARATION ON VIABILITY
issued in accordance with rule 10.2 y the
INTERNATIONAL DEPOSIT AUTHORITY
identified on the following page

NAME AND ADDRESS OF
DEPOSITOR

I – Depositor	II. Identification of the microorganism
Name : **INSTITUT PASTEUR** Address: **Bureau des Brevets et inventions** **25-28 rue du Docteur Roux** **75015 PARIS**	Serial number given by the INTERNATIONAL DEPOSIT AUTHORITY **I-2340** Date of the deposit[1] : **OCTOBER 26, 1999**
II. DECLARATION ON THE VIABILITY	
The viability of the microorganism identified under heading II was controlled On **OCTOBER 27, 1999** [2] At this date the microorganism ■ was viable[3] ☐ was no more viable [3]	

[1-] : Indicate the initial date of the deposit or, if a new deposit or a transfert has been done, the most recent of the relevant dates
[2-] : In the cases referred to in Rule 10.2a)ii) and iii), mention the most recent viability control.
[3-] : Tick the appropriate box

IV. CONDITIONS OF THE VIABILITY CONTROL[4]	
V. INTERNATIONAL DEPOSIT AUTHORITY	
Name : **CNCM** Collection Nationale de Cultures De Microorganismes Address : **INSTITUT PASTEUR** 28, rue du Docteur Roux F-75724 PARIS CEDEX 15	Signature(s) of the person(s) competent to represent the International Deposit Authority or the authorized employee(s): **Yvanne CERISIER** Administrative CNCM Manager [signature] **Georges WAGENER** CNCM Scientific Adviser for Bacteria [signature] Date : **Paris, November 30, 1999**

[4] only fill this part when the control is negative

TRAITE DE BUDAPEST SUR LA RECONNAISSANCE
INTERNATIONALE DU DEPOT DES MICRO-ORGANISMES
AUX FINS DE LA PROCEDURE EN MATIERE DE BREVETS

FORMULE INTERNATIONALE

DESTINATAIRE :

INSTITUT PASTEUR
Bureau des Brevets et Inventions
25-28, rue du Docteur Roux
75015 PARIS

RECEPISSE EN CAS DE DEPOT INITIAL,
délivré en vertu de la règle 7.1 par
l'AUTORITE DE DEPOT INTERNATIONALE
identifiée au bas de cette page

NOM ET ADRESSE
DU DEPOSANT

I. IDENTIFICATION DU MICRO-ORGANISME

Référence d'identification donnée par le DEPOSANT :	Numéro d'ordre attribué par l'AUTORITE DE DEPOT INTERNATIONALE :
β5479	**I - 2339**

II. DESCRIPTION SCIENTIFIQUE ET/OU DESIGNATION TAXONOMIQUE PROPOSEE

Le micro-organisme identifié sous chiffre I était accompagné :

■ d'une description scientifique

■ d'une désignation taxonomique proposée

(Cocher ce qui convient)

III. RECEPTION ET ACCEPTATION

La présente autorité de dépôt internationale accepte le micro-organisme identifié sous chiffre I, qu'elle a reçu le **26 OCTOBRE 1999** (date du dépôt initial) [1]

IV. RECEPTION D'UNE REQUETE EN CONVERSION

La présente autorité de dépôt internationale a reçu le micro-organisme identifié sous chiffre I le (date du dépôt initial)
et a reçu une requête en conversion du dépôt initial en dépôt conforme au Traité de Budapest le (date de réception de la requête en conversion)

V. AUTORITE DE DEPOT INTERNATIONALE

Nom : **CNCM**
Collection Nationale
de Cultures de Microorganismes

Adresse : **INSTITUT PASTEUR**
28, Rue du Docteur Roux
F-75724 PARIS CEDEX 15

Signature(s) de la (des) personne(s) compétente(s) pour représenter l'autorité de dépôt internationale ou de l'(des) employé(s) autorisé(s) : Simona OZDEN
Directeur de la CNCM

Date : **Paris, le 30 novembre 1999**

1 En cas d'application de la règle 6.4.d), cette date est la date à laquelle le statut d'autorité de dépôt internationale a été acquis.

Formule BP/4 (page unique)

BP-4TB.FRA

TRAITE DE BUDAPEST SUR LA RECONNAISSANCE
INTERNATIONALE DU DEPOT DES MICRO-ORGANISMES
AUX FINS DE LA PROCEDURE EN MATIERE DE BREVETS

FORMULE INTERNATIONALE

DESTINATAIRE :

Madame Danielle BERNEMAN,
Bureau des Brevets et Inventions
INSTITUT PASTEUR
25-28, rue du Docteur Roux
75724 PARIS CEDEX 15

NOM ET ADRESSE DE LA PARTIE A LAQUELLE LA DECLARATION SUR LA VIABILITE EST DELIVREE

DECLARATION SUR LA VIABILITE,
délivrée en vertu de la règle 10.2 par
l'AUTORITE DE DEPOT INTERNATIONALE
identifiée à la page suivante

I. DEPOSANT	II. IDENTIFICATION DU MICRO-ORGANISME
Nom : **INSTITUT PASTEUR** Adresse : **Bureau des Brevets et Inventions** **25-28, rue du Docteur Roux** **75015 PARIS**	Numéro d'ordre attribué par l'AUTORITE DE DEPOT INTERNATIONALE : **I - 2339** Date du dépôt ou du transfert [1] : **26 OCTOBRE 1999**

III. DECLARATION SUR LA VIABILITE

La viabilité du micro-organisme identifié sous chiffre II a été contrôlée le **27 OCTOBRE 1999** [2]. A cette date, le micro-organisme

[■] [3] était viable

[ ] [3] n'était plus viable

1 Indiquer la date du dépôt initial ou, si un nouveau dépôt ou un transfert ont été effectués, la plus récente des dates pertinentes (date du nouveau dépôt ou date du transfert).

2 Dans les cas visés à la règle 10.2.a)ii) et iii), mentionner le contrôle de viabilité le plus récent.

3 Cocher la case qui convient.

IV. CONDITIONS DANS LESQUELLES LE CONTROLE DE VIABILITE A ETE EFFECTUE [4]

V. AUTORITE DE DEPOT INTERNATIONALE

Nom :

CNCM
Collection Nationale
de Cultures de Microorganismes

Adresse :

INSTITUT PASTEUR
28, Rue du Docteur Roux
F-75724 PARIS CEDEX 15
FRANCE

Signature(s) de la (des) personne(s) compétente(s) pour représenter l'autorité de dépôt internationale ou de l'(des) employé(s) autorisé(s) :

Simona OZDEN
Directeur de la CNCM

Georges WAGENER
Conseiller Scientifique de la CNCM pour les bactéries

Date : **Paris, le 30 novembre 1999**

4 A remplir si cette information a été demandée et si les résultats du contrôle étaient négatifs.

BUDAPEST TREATY ON THE INTERNATIONAL
RECOGNITION OF THE DEPOSIT OF MICROORANISMS
FOR THE PURPOSES OF PATENT PROCEDURE

INTERNATIONAL SYSTEM

RECIPIENTS:

INSTITUT PASTEUR
Bureau des Brevets et Inventions
25-28, rue du Docteur Roux
75015 PARIS

RECEIPT FOR INITIAL DEPOSIT,
issued in accordance with rule 7.1 by the
INTERNATIONAL DEPOSIT AUTHORITY
identified at the bottom of this page

NAME AND ADDRESS OF
DEPOSITOR

I. IDENTIFICATION OF THE MICROORGANISM	
Identification reference given by the DEPOSITOR **β5479**	Serial number given by the INTERNATIONAL DEPOSIT AUTHORITY **I-2339**

II. SCIENTIFIC DESCRIPTION AND/OR PROPOSED TAXONOMIC DESIGNATION

The microorganism identified under heading I was accompanied:

■ By a scientific description

■ By a proposed taxonomic description

(Check the appropriate box)

III. RECEIPT AND ACCEPTANCE

The present International Deposit Authority accepts the microorganism identified under heading I, which it received on **October 26, 1999** (date of the initial deposit)[1]

IV. RECEIPT OF A REQUEST FOR CONVERSION

The present International Deposit Authority received the microorganism identified under heading I, which it received on (date of the initial deposit)[1]

and received a conversion request of the initial deposit into a deposit which conforms to the Budapest Treaty on (date of the receipt of conversion request)

V. INTERNATIONAL DEPOSIT AUTHORITY

Name : **CNCM** Collection Nationale de Cultures De Microorganismes Address : **INSTITUT PASTEUR** 28, rue du Docteur Roux F-75724 PARIS CEDEX 15	Signature(s) of the person(s) competent to represent the International Deposit Authority or the authorized employee(s): **Simona OZDEN** Director of CNCM [signature] Date : **Paris, November 30, 1999**

1. In the case of application of rule 6.4.d), this date is the date on which the authorizing statute for international deposit was acquired.

BUDAPEST TREATY ON THE INTERNATIONAL
RECOGNITION OF THE DEPOSIT OF MICROORANISMS
FOR THE PURPOSES OF PATENT PROCEDURE

INTERNATIONAL SYSTEM

RECIPIENTS:

INSTITUT PASTEUR
Bureau des Brevets et Inventions
25-28, rue du Docteur Roux
75015 PARIS

DECLARATION ON VIABILITY
issued in accordance with rule 10.2 y the
INTERNATIONAL DEPOSIT AUTHORITY
identified on the following page

NAME AND ADDRESS OF
DEPOSITOR

I – Depositor	II. Identification of the microorganism
Name : **INSTITUT PASTEUR** Address: **Bureau des Brevets et inventions** **25-28 rue du Docteur Roux** **75015 PARIS**	Serial number given by the INTERNATIONAL DEPOSIT AUTHORITY **I-2339** Date of the deposit[1] : **OCTOBER 26, 1999**
II. DECLARATION ON THE VIABILITY	
The viability of the microorganism identified under heading II was controlled On **OCTOBER 27, 1999** [2] At this date the microorganism ■ was viable[3] ☐ was no more viable [3]	

[1-] : Indicate the initial date of the deposit or, if a new deposit or a transfert has been done, the most recent of the relevant dates
[2-] : In the cases referred to in Rule 10.2a)ii) and iii), mention the most recent viability control.
[3-] : Tick the appropriate box

IV. CONDITIONS OF THE VIABILITY CONTROL[4]	
V. INTERNATIONAL DEPOSIT AUTHORITY	
Name : **CNCM** Collection Nationale de Cultures De Microorganismes Address : **INSTITUT PASTEUR** 28, rue du Docteur Roux F-75724 PARIS CEDEX 15	Signature(s) of the person(s) competent to represent the International Deposit Authority or the authorized employee(s): **Yvanne CERISIER** Administrative CNCM Manager [signature] **Georges WAGENER** CNCM Scientific Adviser for Bacteria [signature] Date : **Paris, November 30, 1999**

[4]- only fill this part when the control is negative

TRAITE DE BUDAPEST SUR LA RECONNAISSANCE
INTERNATIONALE DU DEPOT DES MICRO-ORGANISMES
AUX FINS DE LA PROCEDURE EN MATIERE DE BREVETS

FORMULE INTERNATIONALE

DESTINATAIRE :

INSTITUT PASTEUR
Bureau des Brevets et Inventions
25-28, rue du Docteur Roux
75015 PARIS

RECEPISSE EN CAS DE DEPOT INITIAL,
délivré en vertu de la règle 7.1 par
l'AUTORITE DE DEPOT INTERNATIONALE
identifiée au bas de cette page

NOM ET ADRESSE
DU DEPOSANT

I. IDENTIFICATION DU MICRO-ORGANISME

Référence d'identification donnée par le DEPOSANT :	Numéro d'ordre attribué par l'AUTORITE DE DEPOT INTERNATIONALE :
β8144	**I - 2026**

II. DESCRIPTION SCIENTIFIQUE ET/OU DESIGNATION TAXONOMIQUE PROPOSEE

Le micro-organisme identifié sous chiffre I était accompagné :

■ d'une description scientifique

■ d'une désignation taxonomique proposée

(Cocher ce qui convient)

III. RECEPTION ET ACCEPTATION

La présente autorité de dépôt internationale accepte le micro-organisme identifié sous chiffre I, qu'elle a reçu le **25 MAI 1998** (date du dépôt initial) [1]

IV. RECEPTION D'UNE REQUETE EN CONVERSION

La présente autorité de dépôt internationale a reçu le micro-organisme identifié sous chiffre I le (date du dépôt initial)
et a reçu une requête en conversion du dépôt initial en dépôt conforme au Traité de Budapest le (date de réception de la requête en conversion)

V. AUTORITE DE DEPOT INTERNATIONALE

Nom : **CNCM**
Collection Nationale
de Cultures de Microorganismes

Adresse : **INSTITUT PASTEUR**
28, Rue du Docteur Roux
F-75724 PARIS CEDEX 15

Signature(s) de la (des) personne(s) compétente(s) pour représenter l'autorité de dépôt internationale ou de l'(des) employé(s) autorisé(s) : **Mme Y. CERISIER**
Directeur Administratif de la CNCM

Date : **Paris, le 09 juin 1998**

1 En cas d'application de la règle 6.4.d), cette date est la date à laquelle le statut d'autorité de dépôt internationale a été acquis.

Formule BP/4 (page unique)

BP-4TB.FRA

TRAITE DE BUDAPEST SUR LA RECONNAISSANCE
INTERNATIONALE DU DEPOT DES MICRO-ORGANISMES
AUX FINS DE LA PROCEDURE EN MATIERE DE BREVETS

FORMULE INTERNATIONALE

DESTINATAIRE :

Madame D. BERNEMAN,
Bureau des Brevets et Inventions
INSTITUT PASTEUR
25-28, rue du Docteur Roux
75724 PARIS CEDEX 15

NOM ET ADRESSE DE LA PARTIE A LAQUELLE LA DECLARATION SUR LA VIABILITE EST DELIVREE

DECLARATION SUR LA VIABILITE,
délivrée en vertu de la règle 10.2 par
l'AUTORITE DE DEPOT INTERNATIONALE
identifiée à la page suivante

I. DEPOSANT	II. IDENTIFICATION DU MICRO-ORGANISME
Nom : **INSTITUT PASTEUR** Adresse : **Bureau des Brevets et Inventions** **25-28 rue du Docteur Roux** **75015 PARIS**	Numéro d'ordre attribué par l'AUTORITE DE DEPOT INTERNATIONALE : **I - 2026** Date du dépôt ou du transfert [1] : **25 MAI 1998**

III. DECLARATION SUR LA VIABILITE

La viabilité du micro-organisme identifié sous chiffre II a été contrôlée le **26 MAI 1998** [2]. A cette date, le micro-organisme

[■] [3] était viable

[ ] [3] n'était plus viable

1 Indiquer la date du dépôt initial ou, si un nouveau dépôt ou un transfert ont été effectués, la plus récente des dates pertinentes (date du nouveau dépôt ou date du transfert).

2 Dans les cas visés à la règle 10.2.a)ii) et iii), mentionner le contrôle de viabilité le plus récent.

3 Cocher la case qui convient.

IV. CONDITIONS DANS LESQUELLES LE CONTROLE DE VIABILITE A ETE EFFECTUE [4]

V. AUTORITE DE DEPOT INTERNATIONALE	
Nom : **CNCM** Collection Nationale de Cultures de Microorganismes Adresse : **INSTITUT PASTEUR** 28, Rue du Docteur Roux F-75724 PARIS CEDEX 15 FRANCE	Signature(s) de la (des) personne(s) compétente(s) pour représenter l'autorité de dépôt internationale ou de l'(des) employé(s) autorisé(s) : **Yvanne CERISIER** Directeur administratif de la CNCM **Georges WAGENER** Conseiller Scientifique de la CNCM pour les bactéries Date : **Paris, le 09 juin 1998**

4 A remplir si cette information a été demandée et si les résultats du contrôle étaient négatifs.

BUDAPEST TREATY ON THE INTERNATIONAL
RECOGNITION OF THE DEPOSIT OF MICROORANISMS
FOR THE PURPOSES OF PATENT PROCEDURE

INTERNATIONAL SYSTEM

RECIPIENTS:

INSTITUT PASTEUR
Bureau des Brevets et Inventions
25-28, rue du Docteur Roux
75015 PARIS

RECEIPT FOR INITIAL DEPOSIT,
issued in accordance with rule 7.1 by the
INTERNATIONAL DEPOSIT AUTHORITY
identified at the bottom of this page

NAME AND ADDRESS OF
DEPOSITOR

I. IDENTIFICATION OF THE MICROORGANISM	
Identification reference given by the DEPOSITOR **β8144**	Serial number given by the INTERNATIONAL DEPOSIT AUTHORITY **I-2026**

II. SCIENTIFIC DESCRIPTION AND/OR PROPOSED TAXONOMIC DESIGNATION

The microorganism identified under heading I was accompanied:

■ By a scientific description

■ By a proposed taxonomic description

(Check the appropriate box)

III. RECEIPT AND ACCEPTANCE

The present International Deposit Authority accepts the microorganism identified under heading I, which it received on **May 25, 1998** (date of the initial deposit)[1]

IV. RECEIPT OF A REQUEST FOR CONVERSION

The present International Deposit Authority received the microorganism identified under heading I, which it received on (date of the initial deposit)[1]
and received a conversion request of the initial deposit into a deposit which conforms to the Budapest Treaty on (date of the receipt of conversion request)

V. INTERNATIONAL DEPOSIT AUTHORITY

Name : **CNCM** Collection Nationale de Cultures De Microorganismes Address : **INSTITUT PASTEUR** 28, rue du Docteur Roux F-75724 PARIS CEDEX 15	Signature(s) of the person(s) competent to represent the International Deposit Authority or the authorized employee(s): **Mme Y. CERISIER** Administrative director of CNCM [signature] Date : **Paris, June 9, 1998**

1. In the case of application of rule 6.4.d), this date is the date on which the authorizing statute for international deposit was acquired.

BUDAPEST TREATY ON THE INTERNATIONAL
RECOGNITION OF THE DEPOSIT OF MICROORANISMS
FOR THE PURPOSES OF PATENT PROCEDURE

INTERNATIONAL SYSTEM

RECIPIENTS:

INSTITUT PASTEUR
Bureau des Brevets et Inventions
25-28, rue du Docteur Roux
75015 PARIS

DECLARATION ON VIABILITY
issued in accordance with rule 10.2 y the
INTERNATIONAL DEPOSIT AUTHORITY
identified on the following page

NAME AND ADDRESS OF
DEPOSITOR

I – Depositor	II. Identification of the microorganism
Name : **INSTITUT PASTEUR** Address: **Bureau des Brevets et inventions** **25-28 rue du Docteur Roux** **75015 PARIS**	Serial number given by the INTERNATIONAL DEPOSIT AUTHORITY **I-2026** Date of the deposit[1] : **MAY 25, 1998**
II. DECLARATION ON THE VIABILITY	
The viability of the microorganism identified under heading II was controlled On **MAY 26, 1998** [2] At this date the microorganism ■ was viable[3] ☐ was no more viable [3]	

[1-] : Indicate the initial date of the deposit or, if a new deposit or a transfert has been done, the most recent of the relevant dates
[2-] : In the cases referred to in Rule 10.2a)ii) and iii), mention the most recent viability control.
[3-] : Tick the appropriate box

IV. CONDITIONS OF THE VIABILITY CONTROL[4]	
V. INTERNATIONAL DEPOSIT AUTHORITY	
Name : **CNCM** Collection Nationale de Cultures De Microorganismes Address : **INSTITUT PASTEUR** 28, rue du Docteur Roux F-75724 PARIS CEDEX 15	Signature(s) of the person(s) competent to represent the International Deposit Authority or the authorized employee(s): **Yvanne CERISIER** Administrative CNCM Manager [signature] **Georges WAGENER** CNCM Scientific Adviser for Bacteria [signature] Date : **Paris, June 9, 1998**

[4]- only fill this part when the control is negative

TRAITE DE BUDAPEST SUR LA RECONNAISSANCE
INTERNATIONALE DU DEPOT DES MICRO-ORGANISMES
AUX FINS DE LA PROCEDURE EN MATIERE DE BREVETS

FORMULE INTERNATIONALE

DESTINATAIRE :

INSTITUT PASTEUR
Bureau des Brevets et Inventions
25-28, rue du Docteur Roux
75015 PARIS

RECEPISSE EN CAS DE DEPOT INITIAL,
délivré en vertu de la règle 7.1 par
l'AUTORITE DE DEPOT INTERNATIONALE
identifiée au bas de cette page

NOM ET ADRESSE
DU DEPOSANT

I. IDENTIFICATION DU MICRO-ORGANISME

Référence d'identification donnée par le DEPOSANT :	Numéro d'ordre attribué par l'AUTORITE DE DEPOT INTERNATIONALE :
β5366	**I - 2025**

II. DESCRIPTION SCIENTIFIQUE ET/OU DESIGNATION TAXONOMIQUE PROPOSEE

Le micro-organisme identifié sous chiffre I était accompagné :

■ d'une description scientifique

■ d'une désignation taxonomique proposée

(Cocher ce qui convient)

III. RECEPTION ET ACCEPTATION

La présente autorité de dépôt internationale accepte le micro-organisme identifié sous chiffre I, qu'elle a reçu le **25 MAI 1998** (date du dépôt initial) [1]

IV. RECEPTION D'UNE REQUETE EN CONVERSION

La présente autorité de dépôt internationale a reçu le micro-organisme identifié sous chiffre I le (date du dépôt initial)
et a reçu une requête en conversion du dépôt initial en dépôt conforme au Traité de Budapest le (date de réception de la requête en conversion)

V. AUTORITE DE DEPOT INTERNATIONALE

Nom :	Signature(s) de la (des) personne(s) compétente(s) pour représenter l'autorité de dépôt internationale ou de l'(des) employé(s) autorisé(s) :
CNCM Collection Nationale de Cultures de Microorganismes	**Mme Y. CERISIER** Directeur Administratif de la CNCM
Adresse : **INSTITUT PASTEUR** 28, Rue du Docteur Roux F-75724 PARIS CEDEX 15	Date : **Paris, le 09 juin 1998**

1 En cas d'application de la règle 6.4.d), cette date est la date à laquelle le statut d'autorité de dépôt internationale a été acquis.

Formule BP/4 (page unique)

BP-4TB.FRA

TRAITE DE BUDAPEST SUR LA RECONNAISSANCE
INTERNATIONALE DU DEPOT DES MICRO-ORGANISMES
AUX FINS DE LA PROCEDURE EN MATIERE DE BREVETS

FORMULE INTERNATIONALE

DESTINATAIRE :

Madame D. BERNEMAN,
Bureau des Brevets et Inventions
INSTITUT PASTEUR
25-28, rue du Docteur Roux
75724 PARIS CEDEX 15

NOM ET ADRESSE DE LA PARTIE A LAQUELLE LA DECLARATION SUR LA VIABILITE EST DELIVREE

DECLARATION SUR LA VIABILITE,
délivrée en vertu de la règle 10.2 par
l'AUTORITE DE DEPOT INTERNATIONALE
identifiée à la page suivante

I. DEPOSANT	II. IDENTIFICATION DU MICRO-ORGANISME
Nom : **INSTITUT PASTEUR** Adresse : **Bureau des Brevets et Inventions** **25-28 rue du Docteur Roux** **75015 PARIS**	Numéro d'ordre attribué par l'AUTORITE DE DEPOT INTERNATIONALE : **I - 2025** Date du dépôt ou du transfert [1] : **25 MAI 1998**

III. DECLARATION SUR LA VIABILITE

La viabilité du micro-organisme identifié sous chiffre II a été contrôlée le **26 MAI 1998** [2]. A cette date, le micro-organisme

[■] [3] était viable

[ ] [3] n'était plus viable

1 Indiquer la date du dépôt initial ou, si un nouveau dépôt ou un transfert ont été effectués, la plus récente des dates pertinentes (date du nouveau dépôt ou date du transfert).

2 Dans les cas visés à la règle 10.2.a)ii) et iii), mentionner le contrôle de viabilité le plus récent.

3 Cocher la case qui convient.

IV. CONDITIONS DANS LESQUELLES LE CONTROLE DE VIABILITE A ETE EFFECTUE [4]

V. AUTORITE DE DEPOT INTERNATIONALE

Nom : **CNCM**
Collection Nationale
de Cultures de Microorganismes

Adresse : **INSTITUT PASTEUR**
28, Rue du Docteur Roux
F-75724 PARIS CEDEX 15
FRANCE

Signature(s) de la (des) personne(s) compétente(s) pour représenter l'autorité de dépôt internationale ou de l'(des) employé(s) autorisé(s) :

Yvanne CERISIER
Directeur administratif de la CNCM

Georges WAGENER
Conseiller Scientifique de la CNCM pour les bactéries

Date : **Paris, le 09 juin 1998**

4 A remplir si cette information a été demandée et si les résultats du contrôle étaient négatifs.

BUDAPEST TREATY ON THE INTERNATIONAL
RECOGNITION OF THE DEPOSIT OF MICROORANISMS
FOR THE PURPOSES OF PATENT PROCEDURE

INTERNATIONAL SYSTEM

RECIPIENTS:

INSTITUT PASTEUR
Bureau des Brevets et Inventions
25-28, rue du Docteur Roux
75015 PARIS

RECEIPT FOR INITIAL DEPOSIT,
issued in accordance with rule 7.1 by the
INTERNATIONAL DEPOSIT AUTHORITY
identified at the bottom of this page

NAME AND ADDRESS OF
DEPOSITOR

I. IDENTIFICATION OF THE MICROORGANISM	
Identification reference given by the DEPOSITOR **β5366**	Serial number given by the INTERNATIONAL DEPOSIT AUTHORITY **I-2025**

II. SCIENTIFIC DESCRIPTION AND/OR PROPOSED TAXONOMIC DESIGNATION

The microorganism identified under heading I was accompanied:

■ By a scientific description

■ By a proposed taxonomic description

(Check the appropriate box)

III. RECEIPT AND ACCEPTANCE

The present International Deposit Authority accepts the microorganism identified under heading I, which it received on **May 25, 1998** (date of the initial deposit)[1]

IV. RECEIPT OF A REQUEST FOR CONVERSION

The present International Deposit Authority received the microorganism identified under heading I, which it received on (date of the initial deposit)[1]

and received a conversion request of the initial deposit into a deposit which conforms to the Budapest Treaty on (date of the receipt of conversion request)

V. INTERNATIONAL DEPOSIT AUTHORITY

Name : **CNCM** Collection Nationale de Cultures De Microorganismes Address : **INSTITUT PASTEUR** 28, rue du Docteur Roux F-75724 PARIS CEDEX 15	Signature(s) of the person(s) competent to represent the International Deposit Authority or the authorized employee(s): **Mme Y. CERISIER** Administrative director of CNCM [signature] Date : **Paris, June 9, 1998**

1. In the case of application of rule 6.4.d), this date is the date on which the authorizing statute for international deposit was acquired.

BUDAPEST TREATY ON THE INTERNATIONAL
RECOGNITION OF THE DEPOSIT OF MICROORANISMS
FOR THE PURPOSES OF PATENT PROCEDURE

INTERNATIONAL SYSTEM

RECIPIENTS:

INSTITUT PASTEUR
Bureau des Brevets et Inventions
25-28, rue du Docteur Roux
75015 PARIS

DECLARATION ON VIABILITY
issued in accordance with rule 10.2 y the
INTERNATIONAL DEPOSIT AUTHORITY
identified on the following page

NAME AND ADDRESS OF
DEPOSITOR

I – Depositor	II. Identification of the microorganism
Name : **INSTITUT PASTEUR** Address: **Bureau des Brevets et inventions** **25-28 rue du Docteur Roux** **75015 PARIS**	Serial number given by the INTERNATIONAL DEPOSIT AUTHORITY **I-2025** Date of the deposit[1] : **MAY 25, 1998**
II. DECLARATION ON THE VIABILITY	
The viability of the microorganism identified under heading II was controlled On **MAY 26, 1998** [2] At this date the microorganism ■ was viable[3] ☐ was no more viable [3]	

[1-] : Indicate the initial date of the deposit or, if a new deposit or a transfert has been done, the most recent of the relevant dates
[2-] : In the cases referred to in Rule 10.2a)ii) and iii), mention the most recent viability control.
[3-] : Tick the appropriate box

IV. CONDITIONS OF THE VIABILITY CONTROL[4]	
V. INTERNATIONAL DEPOSIT AUTHORITY	
Name : **CNCM** Collection Nationale de Cultures De Microorganismes Address : **INSTITUT PASTEUR** 28, rue du Docteur Roux F-75724 PARIS CEDEX 15	Signature(s) of the person(s) competent to represent the International Deposit Authority or the authorized employee(s): **Yvanne CERISIER** Administrative CNCM Manager [signature] **Georges WAGENER** CNCM Scientific Adviser for Bacteria [signature] Date : **Paris, June 9, 1998**

[4]- only fill this part when the control is negative

TRAITE DE BUDAPEST SUR LA RECONNAISSANCE
INTERNATIONALE DU DEPOT DES MICRO-ORGANISMES
AUX FINS DE LA PROCEDURE EN MATIERE DE BREVETS

FORMULE INTERNATIONALE

DESTINATAIRE :

Madame D. BERNEMAN,
Bureau des Brevets et Inventions
INSTITUT PASTEUR
25-28, rue du Docteur Roux
75724 PARIS CEDEX 15

NOM ET ADRESSE DE LA PARTIE A LAQUELLE LA DECLARATION SUR LA VIABILITE EST DELIVREE

DECLARATION SUR LA VIABILITE,
délivrée en vertu de la règle 10.2 par
l'AUTORITE DE DEPOT INTERNATIONALE
identifiée à la page suivante

I. DEPOSANT	II. IDENTIFICATION DU MICRO-ORGANISME
Nom : **INSTITUT PASTEUR** Adresse : **Bureau des Brevets et Inventions** **25-28 rue du Docteur Roux** **75015 PARIS**	Numéro d'ordre attribué par l'AUTORITE DE DEPOT INTERNATIONALE : **I - 2027** Date du dépôt ou du transfert [1] : **25 MAI 1998**

III. DECLARATION SUR LA VIABILITE

La viabilité du micro-organisme identifié sous chiffre II a été contrôlée le **26 MAI 1998** [2]. A cette date, le micro-organisme

[■] [3] était viable

[ ] [3] n'était plus viable

1 Indiquer la date du dépôt initial ou, si un nouveau dépôt ou un transfert ont été effectués, la plus récente des dates pertinentes (date du nouveau dépôt ou date du transfert).

2 Dans les cas visés à la règle 10.2.a)ii) et iii), mentionner le contrôle de viabilité le plus récent.

3 Cocher la case qui convient.

IV. CONDITIONS DANS LESQUELLES LE CONTROLE DE VIABILITE A ETE EFFECTUE [4]

V. AUTORITE DE DEPOT INTERNATIONALE

Nom : **CNCM**
Collection Nationale
de Cultures de Microorganismes

Adresse : **INSTITUT PASTEUR**
28, Rue du Docteur Roux
F-75724 PARIS CEDEX 15
FRANCE

Signature(s) de la (des) personne(s) compétente(s) pour représenter l'autorité de dépôt internationale ou de l'(des) employé(s) autorisé(s) :

Yvanne CERISIER
Directeur administratif de la CNCM

Georges WAGENER
Conseiller Scientifique de la CNCM pour les bactéries

Date : **Paris, le 09 juin 1998**

4 A remplir si cette information a été demandée et si les résultats du contrôle étaient négatifs.

TRAITE DE BUDAPEST SUR LA RECONNAISSANCE
INTERNATIONALE DU DEPOT DES MICRO-ORGANISMES
AUX FINS DE LA PROCEDURE EN MATIERE DE BREVETS

FORMULE INTERNATIONALE

DESTINATAIRE :

INSTITUT PASTEUR
Bureau des Brevets et Inventions
25-28, rue du Docteur Roux
75015 PARIS

RECEPISSE EN CAS DE DEPOT INITIAL,
délivré en vertu de la règle 7.1 par
l'AUTORITE DE DEPOT INTERNATIONALE
identifiée au bas de cette page

NOM ET ADRESSE
DU DEPOSANT

I. IDENTIFICATION DU MICRO-ORGANISME

Référence d'identification donnée par le DEPOSANT :	Numéro d'ordre attribué par l'AUTORITE DE DEPOT INTERNATIONALE :
β8146	**I - 2027**

II. DESCRIPTION SCIENTIFIQUE ET/OU DESIGNATION TAXONOMIQUE PROPOSEE

Le micro-organisme identifié sous chiffre I était accompagné :

■ d'une description scientifique

■ d'une désignation taxonomique proposée

(Cocher ce qui convient)

III. RECEPTION ET ACCEPTATION

La présente autorité de dépôt internationale accepte le micro-organisme identifié sous chiffre I, qu'elle a reçu le **25 MAI 1998** (date du dépôt initial) [1]

IV. RECEPTION D'UNE REQUETE EN CONVERSION

La présente autorité de dépôt internationale a reçu le micro-organisme identifié sous chiffre I le (date du dépôt initial)
et a reçu une requête en conversion du dépôt initial en dépôt conforme au Traité de Budapest le (date de réception de la requête en conversion)

V. AUTORITE DE DEPOT INTERNATIONALE

Nom : **CNCM**
Collection Nationale
de Cultures de Microorganismes

Adresse : **INSTITUT PASTEUR**
28, Rue du Docteur Roux
F-75724 PARIS CEDEX 15

Signature(s) de la (des) personne(s) compétente(s) pour représenter l'autorité de dépôt internationale ou de l'(des) employé(s) autorisé(s) : **Mme Y. CERISIER**
Directeur Administratif de la CNCM

Date : **Paris, le 09 juin 1998**

1 En cas d'application de la règle 6.4.d), cette date est la date à laquelle le statut d'autorité de dépôt internationale a été acquis.

Formule BP/4 (page unique)

BP-4TB.FRA

BUDAPEST TREATY ON THE INTERNATIONAL
RECOGNITION OF THE DEPOSIT OF MICROORANISMS
FOR THE PURPOSES OF PATENT PROCEDURE

INTERNATIONAL SYSTEM

RECIPIENTS:

INSTITUT PASTEUR
Bureau des Brevets et Inventions
25-28, rue du Docteur Roux
75015 PARIS

RECEIPT FOR INITIAL DEPOSIT,
issued in accordance with rule 7.1 by the
INTERNATIONAL DEPOSIT AUTHORITY
identified at the bottom of this page

NAME AND ADDRESS OF
DEPOSITOR

I. IDENTIFICATION OF THE MICROORGANISM	
Identification reference given by the DEPOSITOR **β8146**	Serial number given by the INTERNATIONAL DEPOSIT AUTHORITY **I-2027**

II. SCIENTIFIC DESCRIPTION AND/OR PROPOSED TAXONOMIC DESIGNATION

The microorganism identified under heading I was accompanied:

■ By a scientific description

■ By a proposed taxonomic description

(Check the appropriate box)

III. RECEIPT AND ACCEPTANCE

The present International Deposit Authority accepts the microorganism identified under heading I, which it received on **May 25, 1998** (date of the initial deposit)[1]

IV. RECEIPT OF A REQUEST FOR CONVERSION

The present International Deposit Authority received the microorganism identified under heading I, which it received on (date of the initial deposit)[1]
and received a conversion request of the initial deposit into a deposit which conforms to the Budapest Treaty on (date of the receipt of conversion request)

V. INTERNATIONAL DEPOSIT AUTHORITY

Name : **CNCM** Collection Nationale de Cultures De Microorganismes Address : **INSTITUT PASTEUR** 28, rue du Docteur Roux F-75724 PARIS CEDEX 15	Signature(s) of the person(s) competent to represent the International Deposit Authority or the authorized employee(s): **Mme Y. CERISIER** Administrative director of CNCM [signature] Date : **Paris, June 9, 1998**

1. In the case of application of rule 6.4.d), this date is the date on which the authorizing statute for international deposit was acquired.

BUDAPEST TREATY ON THE INTERNATIONAL
RECOGNITION OF THE DEPOSIT OF MICROORANISMS
FOR THE PURPOSES OF PATENT PROCEDURE

INTERNATIONAL SYSTEM

RECIPIENTS:

INSTITUT PASTEUR
Bureau des Brevets et Inventions
25-28, rue du Docteur Roux
75015 PARIS

DECLARATION ON VIABILITY
issued in accordance with rule 10.2 y the
INTERNATIONAL DEPOSIT AUTHORITY
identified on the following page

NAME AND ADDRESS OF
DEPOSITOR

I – Depositor	II. Identification of the microorganism
Name : **INSTITUT PASTEUR** Address: **Bureau des Brevets et inventions** **25-28 rue du Docteur Roux** **75015 PARIS**	Serial number given by the INTERNATIONAL DEPOSIT AUTHORITY **I-2027** Date of the deposit[1] : **MAY 25, 1998**
II. DECLARATION ON THE VIABILITY	
The viability of the microorganism identified under heading II was controlled On **MAY 26, 1998** [2] At this date the microorganism ■ was viable[3] ☐ was no more viable [3]	

[1-] : Indicate the initial date of the deposit or, if a new deposit or a transfert has been done, the most recent of the relevant dates
[2-] : In the cases referred to in Rule 10.2a)ii) and iii), mention the most recent viability control.
[3-] : Tick the appropriate box

IV. CONDITIONS OF THE VIABILITY CONTROL[4]	
V. INTERNATIONAL DEPOSIT AUTHORITY	
Name : **CNCM** Collection Nationale de Cultures De Microorganismes Address : **INSTITUT PASTEUR** 28, rue du Docteur Roux F-75724 PARIS CEDEX 15	Signature(s) of the person(s) competent to represent the International Deposit Authority or the authorized employee(s): **Yvanne CERISIER** Administrative CNCM Manager [signature] **Georges WAGENER** CNCM Scientific Adviser for Bacteria [signature] Date : **Paris, June 9, 1998**

[4] only fill this part when the control is negative

TRAITE DE BUDAPEST SUR LA RECONNAISSANCE
INTERNATIONALE DU DEPOT DES MICRO-ORGANISMES
AUX FINS DE LA PROCEDURE EN MATIERE DE BREVETS

FORMULE INTERNATIONALE

DESTINATAIRE :

INSTITUT PASTEUR
Bureau des Brevets et Inventions
25-28, rue du Docteur Roux
75015 PARIS

RECEPISSE EN CAS DE DEPOT INITIAL,
délivré en vertu de la règle 7.1 par
l'AUTORITE DE DEPOT INTERNATIONALE
identifiée au bas de cette page

NOM ET ADRESSE
DU DEPOSANT

I. IDENTIFICATION DU MICRO-ORGANISME

Référence d'identification donnée par le DEPOSANT :	Numéro d'ordre attribué par l'AUTORITE DE DEPOT INTERNATIONALE :
β5486	**I - 2341**

II. DESCRIPTION SCIENTIFIQUE ET/OU DESIGNATION TAXONOMIQUE PROPOSEE

Le micro-organisme identifié sous chiffre I était accompagné :

[■] d'une description scientifique

[■] d'une désignation taxonomique proposée

(Cocher ce qui convient)

III. RECEPTION ET ACCEPTATION

La présente autorité de dépôt internationale accepte le micro-organisme identifié sous chiffre I, qu'elle a reçu le **26 OCTOBRE 1999** (date du dépôt initial) [1]

IV. RECEPTION D'UNE REQUETE EN CONVERSION

La présente autorité de dépôt internationale a reçu le micro-organisme identifié sous chiffre I le (date du dépôt initial)
et a reçu une requête en conversion du dépôt initial en dépôt conforme au Traité de Budapest le (date de réception de la requête en conversion)

V. AUTORITE DE DEPOT INTERNATIONALE

Nom : **CNCM**
Collection Nationale
de Cultures de Microorganismes

Adresse : **INSTITUT PASTEUR**
28, Rue du Docteur Roux
F-75724 PARIS CEDEX 15

Signature(s) de la (des) personne(s) compétente(s) pour représenter l'autorité de dépôt internationale ou de l'(des) employé(s) autorisé(s) : **Simona OZDEN**
Directeur de la CNCM

Date : **Paris, le 30 novembre 1999**

1 En cas d'application de la règle 6.4.d), cette date est la date à laquelle le statut d'autorité de dépôt internationale a été acquis.

Formule BP/4 (page unique)

BP-4TB.FRA

TRAITE DE BUDAPEST SUR LA RECONNAISSANCE
INTERNATIONALE DU DEPOT DES MICRO-ORGANISMES
AUX FINS DE LA PROCEDURE EN MATIERE DE BREVETS

FORMULE INTERNATIONALE

DESTINATAIRE :

Madame Danielle BERNEMAN,
Bureau des Brevets et Inventions
INSTITUT PASTEUR
25-28, rue du Docteur Roux
75724 PARIS CEDEX 15

NOM ET ADRESSE DE LA PARTIE A LAQUELLE LA DECLARATION SUR LA VIABILITE EST DELIVREE

DECLARATION SUR LA VIABILITE,
délivrée en vertu de la règle 10.2 par
l'AUTORITE DE DEPOT INTERNATIONALE
identifiée à la page suivante

I. DEPOSANT	II. IDENTIFICATION DU MICRO-ORGANISME
Nom : **INSTITUT PASTEUR** Adresse : **Bureau des Brevets et Inventions** **25-28, rue du Docteur Roux** **75015 PARIS**	Numéro d'ordre attribué par l'AUTORITE DE DEPOT INTERNATIONALE : **I - 2341** Date du dépôt ou du transfert [1] : **26 OCTOBRE 1999**

III. DECLARATION SUR LA VIABILITE

La viabilité du micro-organisme identifié sous chiffre II a été contrôlée le **27 OCTOBRE 1999** [2]. A cette date, le micro-organisme

[■] [3] était viable

[ ] [3] n'était plus viable

1 Indiquer la date du dépôt initial ou, si un nouveau dépôt ou un transfert ont été effectués, la plus récente des dates pertinentes (date du nouveau dépôt ou date du transfert).

2 Dans les cas visés à la règle 10.2.a)ii) et iii), mentionner le contrôle de viabilité le plus récent.

3 Cocher la case qui convient.

IV. CONDITIONS DANS LESQUELLES LE CONTROLE DE VIABILITE A ETE EFFECTUE [4]

V. AUTORITE DE DEPOT INTERNATIONALE

Nom : **CNCM**
Collection Nationale
de Cultures de Microorganismes

Adresse : **INSTITUT PASTEUR**
28, Rue du Docteur Roux
F-75724 PARIS CEDEX 15
FRANCE

Signature(s) de la (des) personne(s) compétente(s) pour représenter l'autorité de dépôt internationale ou de l'(des) employé(s) autorisé(s) :

Simona OZDEN
Directeur de la CNCM

Georges WAGENER
Conseiller Scientifique de la CNCM pour les bactéries

Date : **Paris, le 30 novembre 1999**

4 A remplir si cette information a été demandée et si les résultats du contrôle étaient négatifs.

BUDAPEST TREATY ON THE INTERNATIONAL
RECOGNITION OF THE DEPOSIT OF MICROORANISMS
FOR THE PURPOSES OF PATENT PROCEDURE

INTERNATIONAL SYSTEM

RECIPIENTS:

INSTITUT PASTEUR
Bureau des Brevets et Inventions
25-28, rue du Docteur Roux
75015 PARIS

RECEIPT FOR INITIAL DEPOSIT,
issued in accordance with rule 7.1 by the
INTERNATIONAL DEPOSIT AUTHORITY
identified at the bottom of this page

NAME AND ADDRESS OF
DEPOSITOR

I. IDENTIFICATION OF THE MICROORGANISM	
Identification reference given by the DEPOSITOR **β5486**	Serial number given by the INTERNATIONAL DEPOSIT AUTHORITY **I-2341**

II. SCIENTIFIC DESCRIPTION AND/OR PROPOSED TAXONOMIC DESIGNATION

The microorganism identified under heading I was accompanied:

■ By a scientific description

■ By a proposed taxonomic description

(Check the appropriate box)

III. RECEIPT AND ACCEPTANCE

The present International Deposit Authority accepts the microorganism identified under heading I, which it received on **October 26, 1999** (date of the initial deposit)[1]

IV. RECEIPT OF A REQUEST FOR CONVERSION

The present International Deposit Authority received the microorganism identified under heading I, which it received on (date of the initial deposit)[1]
and received a conversion request of the initial deposit into a deposit which conforms to the Budapest Treaty on (date of the receipt of conversion request)

V. INTERNATIONAL DEPOSIT AUTHORITY

Name : **CNCM** Collection Nationale de Cultures De Microorganismes Address : **INSTITUT PASTEUR** 28, rue du Docteur Roux F-75724 PARIS CEDEX 15	Signature(s) of the person(s) competent to represent the International Deposit Authority or the authorized employee(s): **Simona OZDEN** Director of CNCM [signature] Date : **Paris, November 30, 1999**

1. In the case of application of rule 6.4.d), this date is the date on which the authorizing statute for international deposit was acquired.

BUDAPEST TREATY ON THE INTERNATIONAL
RECOGNITION OF THE DEPOSIT OF MICROORANISMS
FOR THE PURPOSES OF PATENT PROCEDURE

INTERNATIONAL SYSTEM

RECIPIENTS:

INSTITUT PASTEUR
Bureau des Brevets et Inventions
25-28, rue du Docteur Roux
75015 PARIS

DECLARATION ON VIABILITY
issued in accordance with rule 10.2 y the
INTERNATIONAL DEPOSIT AUTHORITY
identified on the following page

NAME AND ADDRESS OF
DEPOSITOR

I – Depositor	II. Identification of the microorganism
Name : **INSTITUT PASTEUR** Address: **Bureau des Brevets et inventions** **25-28 rue du Docteur Roux** **75015 PARIS**	Serial number given by the INTERNATIONAL DEPOSIT AUTHORITY **I-2341** Date of the deposit[1] : **OCTOBER 26, 1999**
II. DECLARATION ON THE VIABILITY	
The viability of the microorganism identified under heading II was controlled On **OCTOBER 27, 1999** [2] At this date the microorganism ■ was viable[3] ☐ was no more viable [3]	

[1-] : Indicate the initial date of the deposit or, if a new deposit or a transfert has been done, the most recent of the relevant dates
[2-] : In the cases referred to in Rule 10.2a)ii) and iii), mention the most recent viability control.
[3-] : Tick the appropriate box

IV. CONDITIONS OF THE VIABILITY CONTROL[4]	
V. INTERNATIONAL DEPOSIT AUTHORITY	
Name : **CNCM** Collection Nationale de Cultures De Microorganismes Address : **INSTITUT PASTEUR** 28, rue du Docteur Roux F-75724 PARIS CEDEX 15	Signature(s) of the person(s) competent to represent the International Deposit Authority or the authorized employee(s): **Yvanne CERISIER** Administrative CNCM Manager [signature] **Georges WAGENER** CNCM Scientific Adviser for Bacteria [signature] Date : **Paris, November 30, 1999**

[4]- only fill this part when the control is negative

THE JOURNAL OF BIOLOGICAL CHEMISTRY
© 1987 by The American Society of Biological Chemists, Inc.

Vol. 262, No. 15, Issue of May 25, pp. 7189-7194, 1987
Printed in U.S.A.

Structure of the Yeast Valyl-tRNA Synthetase Gene (*VASI*) and the Homology of Its Translated Amino Acid Sequence with *Escherichia coli* Isoleucyl-tRNA Synthetase*

(Received for publication, November 11, 1986)

Xavier Jordana‡, Bruno Chatton, Maria Paz-Weisshaar§¶, Jean-Marie Buhler‖, Fritz Cramer¶, Jean Pierre Ebel, and Franco Fasiolo**

From the Institut de Biologie Moléculaire et Cellulaire, 15 rue René Descartes, 67084 Strasbourg, Cedex, France and the ‖Service de Biochimie, Saclay, F-91191 Gif-sur-Yvette, France

The *VASI* gene encoding the valyl-tRNA synthetase from yeast was isolated and sequenced. The gene-derived amino acid sequence of yeast valyl-tRNA synthetase was found to be 23% homologous to the *Escherichia coli* isoleucyl-tRNA synthetase. This is the highest level of homology reported so far between two distinct aminoacyl-tRNA synthetases and is indicative of an evolutionary relationship between these two molecules. Within these homologous sequences, two functional regions could be recognized: the HIGH region which forms part of the binding site of ATP and the KMSKS region which is recognized as the consensus sequence for the binding of the 3′-end of tRNA (Hountondji, C., Dessen, Ph., and Blanquet, S. (1986) *Biochemie (Paris)* 68, 1071–1078). Secondary structure predictions as well as the presence of both HIGH and KMSKS regions, delineating the nucleotide-binding domain and the COOH-terminal helical domain in aminoacyl-tRNA synthetases of known three-dimensional structure, suggest that the yeast valyl-tRNA synthetase polypeptide chain can be folded into three domains: an NH_2-terminal α-helical region followed by a nucleotide-binding topology and a COOH-terminal domain composed of α-helices which probably carries major sites in tRNA binding.

The aminoacyl-tRNA synthetases are a vastly divergent family of enzymes differing in size and subunit structure but catalyzing the same reaction, the formation of an aminoacyl-tRNA, specific for both the amino acid and the tRNA. The mechanism of the aminoacylation involves the initial rapid formation of an aminoacyladenylate complex followed by the transfer of the aminoacyl moiety to the tRNA. Valyl-tRNA synthetase from yeast is a monomer of M_r 120,000 (Kern *et al.*, 1975) and belongs, together with leucyl- and isoleucyl-tRNA synthetases, to the class of enzymes having the largest polypeptide chain. Activation of a single amino acid by the aminoacyl-tRNA synthetase is, in most cases, very specific. However, valyl- and isoleucyl-tRNA synthetases do not discriminate between closely related amino acids in the adenylate formation step. In neither of these cases, however, is the misactivated amino acid used to form a stable aminoacyl-tRNA. The mechanism of rejection is designated as a proofreading or editing mechanism. The isoleucyl- and valyl-tRNA synthetases are known to hydrolyze the misactivated valyl and threonyl adenylates, respectively (Baldwin and Berg, 1966; Fersht and Kaethner, 1976; Igloi *et al.*, 1977). Knowledge of their structure should be useful in defining structural elements involved in catalysis and/or specificity. The entire primary structure of *Escherichia coli* isoleucyl-tRNA synthetase has been reported (Webster *et al.*, 1984). We present here the isolation and sequence of the *VASI Saccharomyces cerevisiae* gene coding for valyl-tRNA synthetase. Comparison of the translated amino acid sequence with that of isoleucyl-tRNA synthetase from *E. coli* shows the strongest homology ever reported for two distinct aminoacyl-tRNA synthetases.

MATERIALS AND METHODS

Yeast, Bacteria, Plasmids, Gene Libraries, and Growth Media—The yeast genomic bank from *S. cerevisiae* strain X 2180 in phage λgt11 and the host strain Y 1090 (Young and Davis 1983a, 1983b) were kindly provided by Dr. R. Young (Whitehead, MIT). The yeast genomic bank from *S. cerevisiae* strain FL100 in the plasmid vector pFL1 (Chevallier *et al.*, 1980) was a gift from Dr. F. Lacroute (IBMC, Strasbourg, France). The strain FF1.1 (mes1,ura3) was the recipient for yeast transformation (Fasiolo *et al.*, 1981). Parental and transformed yeast strains were grown on YNB (0.67% yeast nitrogen base without amino acids, 2% glucose) supplemented with 100 μg/ml methionine. Transformations of yeast and *E. coli* and preparation of nucleic acids were done using standard procedures.

Enzymes and Reagents—Restriction endonucleases, T4 DNA ligase, and *E. coli* DNA polymerase I (Klenow fragment) were purchased from Boehringer Mannheim. [α-^{32}P]dATP, α-^{35}S-labeled dATP, and ^{125}I were purchased from New England Nuclear.

Antibody Preparation and Plaque Screening—Homogeneous yeast valyl-tRNA synthetase was prepared in our laboratory by Drs. D. Kern and R. Giegé. Rabbits were immunized at 15-day intervals by three subcutaneous injections of 500 μg of enzyme dissolved in 500 μl of 10 mM potassium phosphate buffer (pH 7.4), 150 mM NaCl and emulsified in 500 μl of complete Freund's adjuvant. One week after the last injection, the rabbits were bled, and the immunoglobin fraction was purified from the serum by ammonium sulfate precipitation and DEAE-Sephadex chromatography. Purified antibodies were prepared by chromatography on valyl-tRNA synthetase bound to succinylaminoethyl-Sepharose 4B. Ten nmol of enzyme were coupled to 5 ml of packed gel with *N*-cyclohexyl-*N*′-[β-(*N*-methylmorpholino)ethyl]carbodiimide *p*-toluenesulfonate.

* This work was supported by grants from the Centre National de la Recherche Scientifique. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) J02719.

‡ Supported by a United Nations Education, Science, and Culture Organization long-term fellowship. Permanent address: Faculdad de Medicine, Casilla 6667, Santiago 7, Chile.

§ Supported by a European Molecular Biology Organization short-term fellowship.

¶ Permanent address: Max-Planck-Institut fur Experimentelle Medizin, Abteilung Chemie, Hermann-Rein-Strasse 3, D-3400 Goettingen, West Germany.

** To whom correspondence should be addressed.

Screening of the λgt11 genomic library was carried out essentially as described by Young and Davis (1983b) using affinity-purified antibodies at a concentration of 5–10 μg/ml and ^{125}I-protein A (50 μCi/μg) at 1 μCi/ml. Positive plaques were purified by four additional cycles of screening.

Hybridization Procedures—DNA probes were purified by gel electrophoresis or sucrose gradient centrifugation from phage λgt11 or recombinant plasmids digested with the appropriate restriction enzymes. They were labeled by nick translation as described by Maniatis *et al.* (1982). DNA probes cloned in M13 phage were labeled by chain extension using the Klenow fragment of *E. coli* DNA polymerase I and [α-^{32}P]dATP. The yeast genomic bank in vector pFL1 was screened by the high density colony-screening procedure described by Hanahan and Meselson (1983). Positives clones were purified by two additional cycles of screening. Southern blot hybridizations were carried out according to the procedures described by Maniatis *et al.* (1982).

Determination of Enzymatic Activities—Cytoplasmic valyl-tRNA synthetase was tested in crude extracts obtained by mechanical breakage with glass beads of exponentially growing cells. Protein concentration was estimated according to Bradford (1976).

The enzyme was tested using unfractionated yeast cytoplasmic tRNA under the following conditions: 144 mM Tris-HCl (pH 7.8), 5 mM dithiothreitol, 10 mM ATP, 2 mM $MgCl_2$, 0.1 mM [^{14}C]valine (25,000 cpm/nmol), 6 mg/ml yeast tRNA, and various amounts of crude extracts. The reaction mixture was 200 μl; and at various time intervals, 40-μl aliquots were spotted onto Whatman paper discs and quenched with 5% trichloroacetic. The precipitated aminoacylated tRNA was subjected to scintillation counting.

Western Blot—Protein samples were run on 10% polyacrylamide gels in the presence of 0.1% sodium dodecyl sulfate (Laemmli, 1970). Conditions for the transfer of proteins to nitrocellulose membranes were as described in the Schleicher & Schuell manual (No. 2). The protein band corresponding to valyl-tRNA synthetase was detected as described above using affinity-purified antibodies (5–10 μg/ml and ^{125}I-protein A (0.1 μCi/ml).

DNA Sequence Analysis—The dideoxy-DNA sequencing method of Sanger *et al.* (1977) was used. *Eco*RI and *Sal*I digestions of pVASI recombinant generated fragments of 1.6, 1.2, and 1.3 kb, respectively. These DNA fragments were isolated and digested with *Alu*I, *Hae*III, *Taq*I, and *Sau*3A. The resulting subfragments as well as the original fragments were cloned into suitable M13mp8 and M13mp9 vectors (Vieira and Messing, 1982).

Computer Analysis of Amino Acid Sequences—Amino acid sequences were analyzed with programs of the University of Wisconsin Genetics Computer Group edited by Dereveux and Haeberli[1] to locate sequences patterns: "Best fit" to align two sequences; "Gap" to find the optimal alignment for two sequences by adding gaps in either one to achieve the maximum number of matches; "Dotplot" and "Pepplot" to visualize the homology between two sequences; and "Choufas" to perform prediction of secondary structures.

RESULTS

Cloning of the VASI Gene—We have screened a yeast DNA library using the expression vector λgt11 which contains random genomic fragments in the unique *Eco*RI site (Young and Davis, 1983a, 1983b). Ten putative positive clones were obtained and further purified by three successive rounds of antibody screening at low plaque density after which only one clone remained positive. Yeast DNA inserted into the λgt11 recombinant is 2.5 kb,[2] whereas the minimum expected length of the message for a protein of M_r 120,000 (Kern *et al.*, 1975) is about 3.5 kb. In order to isolate the complete gene coding for valyl-tRNA synthetase, we have screened the pFL1 yeast DNA library (Chevallier *et al.*, 1980) using the yeast *Eco*RI fragment from the λgt11 recombinant as hybridization probe. Only three clones (pVASI-1, -2, and -3) were purified, and their overlapping inserts were mapped with a number of restriction enzymes. Southern blot hybridization analysis of yeast nuclear DNA gave an identical genomic map for the two *Eco*RI and *Hin*dIII sites (Fig. 1).

To demonstrate that the cloned gene codes for valyl-tRNA synthetase, we expressed the various clones in yeast to give catalytically active valyl-tRNA, synthetase. The activity in the crude extracts of the yeast transformants (pVASI-1 and -2) was approximatively 10 times higher than the basal level of enzyme in the recipient strain. In order to verify that the activity was associated with a full-length protein in the overproducing strains, proteins from a crude cytoplasmic extract were separated by electrophoresis on sodium dodecyl sulfate-polyacrylamide gels and transferred to nitrocellulose, and valyl-tRNA synthetase was detected using the specific cytoplasmic valyl-tRNA synthetase antibodies and ^{125}I-labeled protein A. The results of the Western blot analyses are shown in Fig. 2. A protein band which co-migrated with the purified cytoplasmic valyl-tRNA synthetase was detected in the crude extract of the recipient strain (*lane 2*). The concentration of this protein was increased (*lanes 3* and *4*) in yeast transformants harboring the *VASI* gene on a multicopy plasmid (pVASI-1 and -2). The level of valyl-tRNA synthetase in the transformant corresponding to clone pVASI-3 was again similar to the basal level of the recipient strain and was probably due to lack of the 5′-upstream promoter sequences.

Determination of the Nucleotide Sequence of the VASI Gene—We have determined 80% of the entire sequence on both strands, and on one strand, the remaining 20%. All restriction endonuclease sites used for generating M13 clones were overlapped. This strategy enabled us to localize a 78-base pair *Eco*RI fragment between the large 1.6- and 1.2-kb *Eco*RI subfragments. A long open reading frame of 3,312 nucleotides was found only on one strand (Fig. 3). The trans-

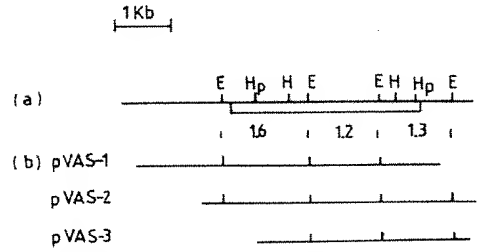


FIG. 1. **Restriction map in the *VASI* genomic region.** *a*, the restriction map was determined by Southern analysis using yeast genomic DNA. The *box* indicates the extent of the *VASI* coding region. The *numbers* refer to the size (in kilobases) of the *Eco*RI fragments. *b*, clones obtained from the screening of the pFL1 yeast DNA library are designated pVASI-1 to -3. *E*, *Eco*RI; *Hin*dIII; *Hp*, *Hpa*I. pVASI clones were aligned with respect to *Eco*RI fragments.

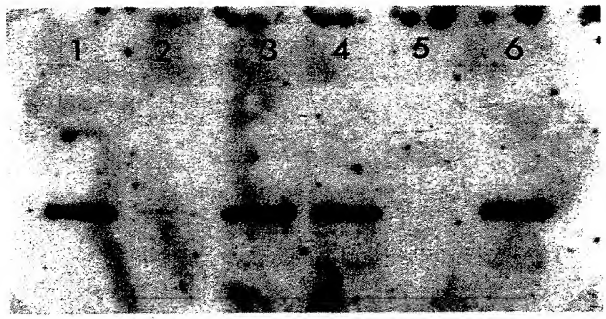
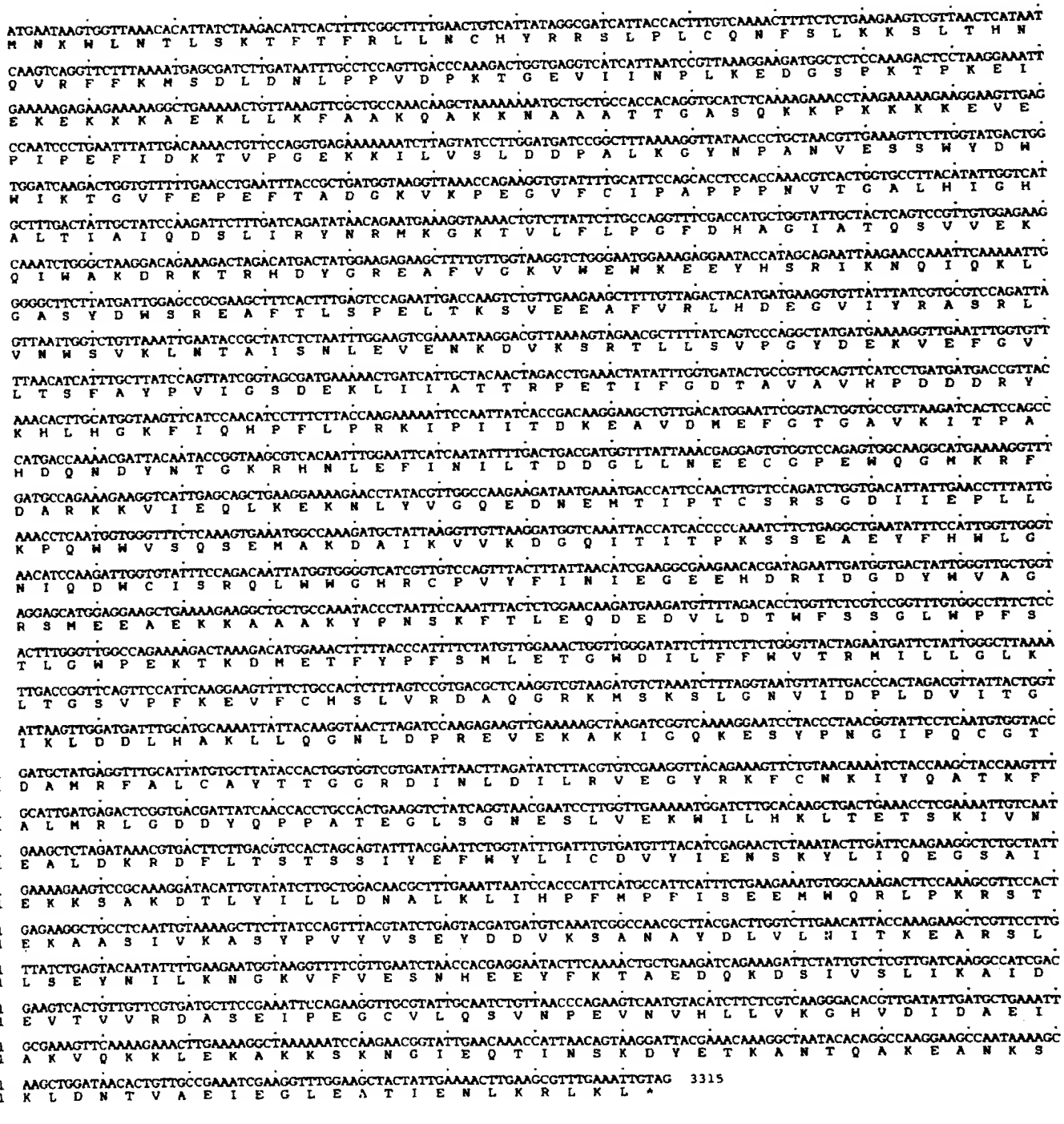


FIG. 2. **Western blot of valyl-tRNA synthetase in crude extracts from recipient and yeast transformants.** *Lanes 1* and *6*, 100 ng of purified cytoplasmic valyl-tRNA synthetase; *lanes 2–5*, 30 μg of cytosol protein from recipient (*lane 2*), transformant pVASI-1 (*lane 3*), transformant pVASI-2 (*lane 4*), and transformant pVASI-3 (*lane 5*).

[1] Dereveux, J., and Haeberli, P. (1983) Program Library of the University of Wisconsin Genetics Computer Group, Madison, WI.

[2] The abbreviation used is: kb, kilobase.



lated amino acid sequence from the first in-phase methionine codon includes 1,104 amino acid residues, yielding a protein of M_r 125,000, in good agreement with the M_r measured for the purified protein. Attempts to define the NH_2-terminal peptide of the protein were unsuccessful due to a blocked NH_2 terminus.

DISCUSSION

Sequence homologies among different aminoacyl-tRNA synthetases, with the exception of those specific for the same amino acid in different organisms, are rare or nonexistent. Similarities of the three-dimensional level of these enzymes, however, are expected to be much greater due to structural constraints imposed on the binding of tRNA which probably shares the same tertiary conformation (Moras *et al.*, 1980) and to the necessity of bringing the adenylate site close to the terminal adenosine site of tRNA in order to achieve the chemical acylation step. Since the ATP and the 3′-CCA arm of tRNA are common to all aminoacyl-tRNA synthetases, it is reasonable to assume that identical or at least functionally

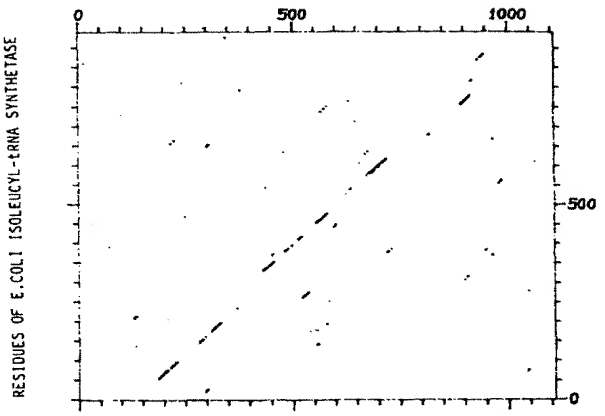
equivalent residues are present in many aminoacyl-tRNA synthetases. Hence, a comparison of primary sequences can be useful to identify important binding and/or catalytic residues. A classical example derives from a structural comparison of *E. coli* methionyl-tRNA synthetase and *Bacillus stearothermophilus* tyrosyl-tRNA synthetase (Blow *et al.*, 1983). The three-dimensional structures of both enzymes indicate folding of the NH_2-terminal regions into similar and characteristic nucleotide-binding domains, although there is only a short stretch of amino acid sequence homology. In particular, 1 cysteine and 2 histidine residues occupy identical positions in the two tertiary structures (Barker and Winter, 1982; Blow *et al.*, 1983). These conserved residues are involved in the binding and catalysis of adenylate formation as demonstrated by site-directed mutagenesis experiments (Winter *et al.*, 1982; Leatherbarrow *et al.*, 1985).

The NH_2-terminal region of *E. coli* isoleucyl-tRNA synthetase shows a sequence homology of 11 consecutive amino acids with the corresponding region of *E. coli* methionyl-tRNA synthetase which allowed the authors (Webster *et al.*, 1984) to conclude that isoleucyl-tRNA synthetase is similarly folded in an alternating β/α structure. The perfect peptide match includes the consensus HIGH region involved in ATP binding (see below).

We have compared the deduced amino acid sequences of yeast valyl-tRNA synthetase and *E. coli* isoleucyl-tRNA synthetase. Residues 177–726 of the yeast enzyme could be aligned with residues 50–618 of the bacterial enzyme (Fig. 4). Fig. 4 shows four short perfect matches of 5–13 conserved residues at the following peptide positions in the yeast sequence: 196–200, 431–435, 564–570, and 700–712. The overall homology is 23%.[3] Two functional regions can be recognized within this homology; one at the ATP-binding site and the other at the probable CCA-binding site of tRNA.

Homology at the ATP-binding Site—Fig. 5 compares the homologies centered around the HIGH region of tyrosyl-tRNA synthetase from *B. stearothermophilus*, methionyl- and isoleucyl-tRNA synthetases from *E. coli*, and the methionyl- and valyl-tRNA synthetases from *S. cerevisiae*. The importance of the HIGH region in ATP binding and catalysis has become apparent from the studies of Fersht *et al.* (1984). This region is in the NH_2-terminal portion of the bacterial enzymes mentioned above, as is the case for the majority of prokaryotic tRNA synthetases; whereas we located the HIGH sequence in both yeast methionyl- and valyl-tRNA synthetases to approximately 200 amino acid residues from the NH_2-terminus. That this region corresponds to the ATP-binding site in yeast valyl-tRNA synthetase can be deduced by analogy with similar positions of the folded α/β topology in yeast methionyl-tRNA synthetase (Walter *et al.*, 1983). Thus, the two yeast enzymes bear an NH_2-terminal chain extension with respect to the mononucleotide binding fold. In yeast valyl-tRNA synthetase, this NH_2-terminal extension is mainly an α-helical region as deduced from predicted secondary structures.

Homology at the CCA-binding Site of tRNA—Covalent labeling of methionyl-tRNA synthetase from *E. coli* with 2′,3′-dialdehyde $tRNA_{ox}^{Met}$ has led to the identification of a peptide encompassing Lys-335 (Hountondji and Blanquet, 1985). Although the exact position of this lysine residue in the crystal structure has not yet been located, it is part of the COOH-terminal helical domain of the synthetase (see Brunie *et al.*



B

```
178 KPEGVFCIPAPPPNVTGALHIGHALTIAIQDSLIRYNRMKGKTVLFLPGF 227
 46 KGKKTFILHDGPPYANGSIHIGHSVNKILKDIIVKSKGLSGYDSPYVPGW 95
228 DHAGIATQSVVEKQIWAKDRKTRHDYGREAFVGKVWEWKEEYHSRIKNQI 277
 96 DCHGLPIELKVE....QEYGKPGEKFTAAEFRAKCREYAATQVDGQRKDF 141
278 QKLGASYDWSREAFTLSPELTKSVEEAFVRLHDEGVIYRASRLVNWSVKL 327
142 IRLGVLGDWSHPYLTMDFKTEANIIRALGKIIGNGHLHKGAKPVHWCVDC 191
328 NTAISNLEVENKDVKSRTLLSVPGYDEKVEFGVLTSFAYPVIGSDEKLII 377
192 RSALAEAEVEYYDKTS..PSIDVAFQAVDQDALKAKFAVSNVNGPISLVI 239
378 ATTRPETIFGDTAVAVH..........PDDDRYKHLHGKFIQHPFLPR 415
240 WTTRRGLCLPTAQSLLHQISTNRWNQIDGQAVILAKDLVESVMQRIGVTD 289
416 KIPIITDKEA.....................VDMEFGTGAVKITPAHD 442
290 YTILGTVKGADVELLRFTHPFWGFDVPAILGDHVTLDAGTGAVHTAPGHG 339
443 QNDYNTGKRHNLEFINILTDDGLLNEECGPEWQGMKRFDARKKVIEQLKE 492
340 PDDYVIGQKYGLETANPVGPDGTYLPGTYPTLDGVNVFKANDIVVALLQE 389
493 KNLYVGQEDWEMTIPTCSRSGDIIEPLLKPQWWVSQSEMAKDAIKVVKDG 542
390 KGALLHVEKMQHSVPCCWRHKTPIIFRATPQWFVSMDQKGLRAQSLKEIK 439
543 QITITPKSSEAEYFHWLGNIQDWCISRQLWWGHRCPVYFINIEGEEHDRI 592
440 GVQWIPDWGQARIESMVANRPDWCISRQRTWGVPWSLFVHKDTEELWPRT 489
593 DGDVWVAGRSWEEAEKKAAAKYPNSKFTLEQDEDVLDTWFSSGLW..PFS 640
490 LELMEEVAKRVEVDGIQAWWDLDAKEILGDEADQYVKVPDTLDVWFDSGS 539
641 TLGWPEKTKDMEIFYPFSMLETGWDILFFWVIRMILLGLKLTGSVPFKEV 690
540 THSSVVDVRPEFAGHAADMYLEGSDQHRGWFMSSLMISTAMKGKAPYCQV 589
691 FCHSLVRDAQGRKMSKSLGNVIDPLDVITGIKLDDLHAKLLQGNLDPREV 740
590 LTHGFTVDGQGRKMSKSIGNTVSPQDVMNKLGADILRLWVASTDYTGQMA 639
741 EKAKIGQKESYPNGIPQCGTDAMRFALCAYTTGGRDINLDILRVEGYRKF 790
640 VSDEI.LKRAADSYRRIRNTARFLLANLNGFDPAKDMVKPEEMVVLDRWA 688
791 CNKIYQATKFALMRLGDDYQPPATEGLSGNESLVEKWILHKLTETSKIVN 840
689 VGCAKAAQEDILKAYEAYDFHEVVQRLMRFCSVEMVSFYLDIIKDRQYTP 738
841 EALDKRDFLTSTSSIYEFWYLICDVYIENSKYLIQEGSAIEKKSAKDTLY 890
739 KRTVWARRSCQTALYHIAEALVRWMAPILSFTADEVWGYLPGEREKYVLT 788
891 ILLDWALKLIHPFMPFISEEMWQRLPKR.STEKAASIVKASYPVYVSEYD 939
789 GEWYEGLFGLADSEAMNDAFWDELLKVRGEVNKVIEQARADNKVGGSLEA 838
940 DVKSANAYDLVLNITKEARSLLSEYNILKNGKVFVESNHEEYFKTAEDQK 989
839 AVTLYAEPELSAKLTALGDELRFVLLTSDRRYVADYNDAPADAQQSEVLK 888
990 DSIVSLIKAIDEVTVVRDASEIPEGCVLQSVNPEVNVHLLVKG 1032
889 GLKVALSKAEGEKCPRCWHYTQDVGKVAEHAEICGRCVSNVAG 931
```

FIG. 4. **Homology between the amino acid sequences of yeast valyl-tRNA synthetase and *E. coli* isoleucyl-tRNA synthetase.** The comparisons shown in both *A* and *B* use programs from the University of Wisconsin Genetics Computer Group. The *E. coli* sequence is from Webster *et al.* (1984). The comparison in *A* uses the Dot Matrix program. Average score values were calculated for pairs of 25-amino acid segments using the mutation matrix of Staden (1982). If the average score value was equal to or greater than 25, a dot was printed at the corresponding position of the matrix. In *B*, the

[3] The sequence of the *E. coli* gene coding for valyl-tRNA synthetase was sent to us before publication by Dr. R. Leberman (LEBM, Grenoble, France) and co-workers. It turned out that the protein sequence was 45% homologous to the yeast enzyme and 23% homologous to the *E. coli* isoleucyl-tRNA synthetase.

(a)	TyrRSbs	33	L Y C G F D P T A D S L H I G H L A T I	52
(b)	MetRS	9	V T C A L P Y A N G S I H L G H M L E H	28
(c)	MetRSsc	200	I T S A L P Y V N N V P H L G N I I G S	219
(d)	IleRS	53	L H D G P P Y A N G S I H I G H S V N K	72
(e)	ValRSsc	183	I P A P P P N V T G A L H I G H A L T.I	204

FIG. 5. **Alignment of the amino acid sequences from the HIGH regions.** The numbering indicates the distance from the NH_2 terminus. The *letters* in *parentheses* indicate the reference of the sequence: *a*, Winter *et al.*, 1983; *b*, Barker and Winter, 1982; *c*, Walter *et al.*, 1983; *d*, Webster *et al.*, 1984; and *e*, this work. *TyrRSbs*, *B. stearothermophilus* tyrosyl-tRNA synthetase; *MetRS*, *E. coli* methionyl-tRNA synthetase; *MetRSsc*, *S. cerevisiae* methionyl-tRNA synthetase; *IleRS*, *E. coli* isoleucyl-tRNA synthetase; *ValRSsc*, *S. cerevisiae* valyl-tRNA synthetase.

(a)	TyrRS	223	T V P L I T K A D G T K F G K L - T	238
(b)	MetRS	329	N G A K M S K S R G T - F I K A S	344
(c)	MetRSsc	522	E N G K F S K S R G V	532
(d)	IleRS	599	Q G R K M S K S I G N T V S P Q D	615
(e)	ValRSsc	700	Q G R K M S K S L G N V I D P L D	716

FIG. 6. **Alignment of the amino acid sequences around the KMSKS regions.** The origins of the sequences are indicated by the same nomenclature used in Fig. 5. The numbering indicates the distance from the NH_2 terminus. The references are as follows: *a*, Barker *et al.*, 1982a; *b*, Barker *et al.*, 1982b; *c*, Walter *et al.*, 1983; *d*, Webster *et al.*, 1984; and *e*, this work.

cited in Hountondji and Blanquet, 1985). The functional importance of the tRNA synthetase region corresponding to Lys-335 is further supported by labeling of a similar sequence in *E. coli* tyrosyl-tRNA synthetase using [^{14}C]tRNA$_{Ot}^{Tyr}$ (Hountoudji *et al.*, 1986a). The labeled lysines at positions 229, 234, and 237 belong to a sequence which is highly conserved in *B. stearothermophilus* tyrosyl-tRNA synthetase (Winter *et al.*, 1983), and their spatial positions were deduced by analogy with the known three-dimensional structure of the homologous *B. stearothermophilus* enzyme (Hountondji *et al.*, 1986a). These lysines are part of the COOH-terminal domain, in the middle of the β-turn joining the last β-strand of the nucleotide domain to the first helix of the helical domain (Bhat *et al.*, 1982), hence in close contact with the adenylate site. The corresponding lysines in the *B. stearothermophilus* enzyme are located at positions 225, 230, and 233. Bedouelle and Winter (1986) could demonstrate that mutations at Lys-151, Arg-207, and Lys-208 also affect the binding of the 3′-end of tRNA. These results are not conflicting since the residues lie on the rim of the tyrosyl adenylate pocket (Bedouelle and Winter, 1986). Fig. 6 indicates the alignment of the reactive lysines characterized in methionyl- and tyrosyl-tRNA synthetases from *E. coli* with similar regions of *E. coli* isoleucyl-tRNA synthetases and the yeast valyl- and methionyl-tRNA synthetases. A more complete overview of similar regions of other aminoacyl-tRNA synthetases is presented by Hountondji *et al.* (1986b). This comparison indicates the presence of the relevant KMSKS sequence which probably represents the consensus sequence of the binding region of the 3′-end of tRNA. This sequence is also conserved in the primary structures of the three homologous tryptophanyl-tRNA synthetases of prokaryotic and eukaryotic origins (Myers and Tzagoloff, 1985). Fig. 6 emphasizes the fact that the KMSKS region is conserved in the valyl/isoleucyl-tRNA synthetase pair within the perfect match of 13 amino acid residues.

E. coli methionyl-tRNA synthetase is structurally similar to the *B. stearothermophilus* tyrosyl-tRNA synthetase (Zelwer *et al.*, 1982; Bhat *et al.*, 1982). They are biglobular enzymes composed of an NH_2-terminal α/β domain (approximately 200 residues) connected through a long loop to an α-helical rich COOH-terminal domain. The latter is responsible for tRNA binding. This is seen in the tyrosyl-tRNA synthetase by protein engineering (Bedouelle and Winter, 1986) and by creating a deletion in the corresponding gene so as to remove 100 residues in the COOH-terminal region, yielding a truncated enzyme able to activate the amino acid but unable to carry the aminoacylation step (Waye *et al.*, 1983). Since the CCA arm is close to the adenylate site, the geometry of the tRNA molecule imposes interaction of the anticodon stem with the COOH-terminal end of the enzyme at a distance of 75 Å from the 3′-end of tRNA. Protein engineering confirms that two separated clusters of basic residues Arg-368-Arg-371 and Arg-407-Arg-408-Lys-410-Lys-411 at the end of the polypeptide chain of tyrosyl-tRNA synthetase from *B. stearothermophilus* (Ala-419) interact with the anticodon stem. The correlation between each catalytic function of the tRNA synthetase and the existence of a distinct structural domain was postulated earlier (Jasin *et al.*, 1983) and can also be deduced in the case of yeast valyl-tRNA synthetase from the presence of the relevant amino acid sequences mentioned above.

Amino acid residue 200 would grossly define the beginning of the nucleotide binding fold, and the KMSKS region at residue 702 would locate the beginning of the α-helical COOH-terminal domain. In that respect, we notice the presence of an α-helical region in the COOH-terminal part of the enzyme according to secondary structure prediction. Furthermore, the presence of a cluster of lysines from residues 952 to 1054 may represent potential contact points with tRNA^{Val} anticodon stem.

We asked the question whether the homology between isoleucyl- and valyl-tRNA synthetases is indicative of a functional relationship (the isoleucyl-tRNA synthetase misactivates valine) or of an evolutionary relationship between these two molecules. Twenty percent sequence homology, as reported here for valyl- and isoleucyl-tRNA synthetases, has only been observed to date for those enzymes specific for the same amino acid but isolated from different organisms, *i.e.* the methionyl-tRNA synthetase pair from *E. coli* and yeast; the homology is even larger for the threonyl/tryptophanyl-tRNA synthetase pairs from *E. coli* and yeast and the tyrosyl-tRNA synthetase pair from *E. coli* and *B. stearothermophilus* (cited by Hountondji *et al.*, 1986b). In contrast, no homology has been identified between two distinct aminoacyl-tRNA synthetases specific for a given amino acid, except for the functional regions mentioned above. In particular, there is no homology between yeast valyl-tRNA synthetase and *E. coli* threonyl-tRNA synthetase which could have explained the misactivation of the isosteric valine-threonine amino acids. Rather, the homology between valyl- and isoleucyl-tRNA synthetases reported in this work suggests an evolution from a common ancestral gene.

amino acid sequences of the two proteins were aligned by the Align program. Breaks in the sequence are shown as *dots*, and identities between amino acid residues are shown by *vertical lines*. The number at the beginning of each line corresponds to the number of the residue in the yeast protein sequences (*upper line*) and the *E. coli* protein sequence (*lower line*).

Acknowledgments—We thank Professor Y. Boulanger and Dr. P. Remy for critically reviewing this work and helpful comments and Dr. Ph. Walter for his help in the computer analysis. We acknowledge the skillful technical assistance of G. Nussbaum.

REFERENCES

Baldwin, A. N., and Berg, P. (1966) *J. Biol. Chem.* **241,** 839–842

Barker, D. G., and Winter, G. (1982) *FEBS Lett.* **145,** 191–193

Barker, D. G., Bruton, C. J., and Winter, G. (1982a) FEBS Lett. **150,** 419–423

Barker, D. G., Ebel, J. P., Jakes, R., and Bruton, C. J. (1982b) *Eur. J. Biochem.* **127,** 449–457

Bedouelle, H., and Winter, G. (1986) *Nature* **320,** 371–373

Bhat, T. N., Blow, D. M., Brick, P., and Nyborg, J. (1982) *J. Mol. Biol.* **158,** 699–709

Blow, D. M., Bhat, T. N., Metcalfe, A., Risler, J. L., Brunie, S., and Zelwer, C. (1983) *J. Mol. Biol.* **171,** 571–576

Bradford, M. M. (1976) *Anal. Biochem.* **72,** 248–254

Chevallier, M. R., Bloch, J. C., and Lacroute, F. (1980) *Gene (Amst.)* **11,** 11–19

Fasiolo, F., Bonnet, J., and Lacroute, F. (1981) *J. Biol. Chem.* **256,** 2324–2328

Fersht, A. R., and Kaethner, M. M. (1976) *Biochemistry* **15,** 3342–3345

Fersht, A. R., Shi, J. P., Wilkinson, A. J., Blow, D. M., Carter, P., Waye, M. M. Y., and Winter, G. P. (1984) *Angew. Chem. Int. Ed. Engl.* **23,** 467–473

Hanahan, D., and Meselson, M. (1983) *Methods Enzymol.* **100,** 333–342

Hountondji, C., and Blanquet, S. (1985) *Biochemistry* **24,** 47–52

Hountondji, C., Lederer, F., and Blanquet, S. (1986a) *Biochemistry* **24,** 1175–1180

Hountondji, C., Dessen, Ph., and Blanquet, S. (1986b) *Biochimie (Paris)* **68,** 1071–1078

Igloi, G. L., von der Haar, F., and Cramer, F. (1977) *Biochemistry* **16,** 1696–1700

Jasin, M., Regan, L., and Schimmel, P. (1983) *Nature* **306,** 441–447

Kern, D., Giegé, R., Robbe-Saul, S., Boulanger, Y., and Ebel, J. P. (1975) *Biochimie (Paris)* **57,** 1167–1176

Laemmli, U. K. (1970) *Nature* **227,** 680–685

Leatherbarrow, R. J., Fersht, A. R., and Winter, G. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82,** 7840–7844

Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual,* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Moras, D., Comarmond, M. B., Fischer, J., Weiss, R., Thierry, J. C., Ebel, J. P., and Giegé, R. (1980) *Nature* **288,** 669–674

Myers, A. M., and Tzagoloff, A. (1985) *J. Biol. Chem.* **260,** 15371–15377

Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5463–5467

Staden, M. (1982) *Nucleic Acids Res.* **10,** 2951–2961

Vieira, J., and Messing, J. (1982) *Gene (Amst.)* **19,** 259–268

Walter, P., Gangloff, J., Bonnet, J., Boulanger, Y., Ebel, J. P., and Fasiolo, F. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80,** 2437–2441

Waye, M. M. Y., Winter, G., Wilkinson, A. J., and Fersht, A. R. (1983) *EMBO J.* **2,** 1827–1829

Webster, T., Tsar, H., Kula, M., Mackie, G. A., and Schimmel, P. (1984) *Science* **226,** 1315–1317

Winter, G., Fersht, A. R., Wilkinson, A. J., Zoller, M., and Smith, M. (1982) *Nature* **299,** 756–758

Winter, G., Koch, G. L. E., Hartley, B. S., and Barker, D. G. (1983) *Eur. J. Biochem.* **132,** 383–387

Young, R. A., and Davis, R. W. (1983a) *Proc. Natl. Acad. Sci. U. S. A.* **80,** 1194–1198

Young, R. A., and Davis, R. W. (1983b) *Science* **222,** 778–782

Zelwer, C., Risler, J. L., and Brunie, S. (1982) *J. Mol. Biol.* **155,** 63–81

THE JOURNAL OF BIOLOGICAL CHEMISTRY
© 1988 by The American Society for Biochemistry and Molecular Biology, Inc.

Vol. 263, No. 2, Issue of January 15, pp. 868-877, 1988
Printed in U.S.A.

Valyl-tRNA Synthetase Gene of *Escherichia coli* K12

PRIMARY STRUCTURE AND HOMOLOGY WITHIN A FAMILY OF AMINOACYL-TRNA SYNTHETASES*

(Received for publication, June 8, 1987)

J. Denis Heck‡ and G. Wesley Hatfield§

From the Department of Microbiology and Molecular Genetics, California College of Medicine, University of California, Irvine, Irvine, California 92717

The DNA nucleotide sequence of the *valS* gene encoding valyl-tRNA synthetase of *Escherichia coli* has been determined. The deduced primary structure of valyl-tRNA synthetase was compared to the primary sequences of the known aminoacyl-tRNA synthetases of yeast and bacteria. Significant homology was detected between valyl-tRNA synthetase of *E. coli* and other known branched-chain aminoacyl-tRNA synthetases. In pairwise comparisons the highest level of homology was detected between the homologous valyl-tRNA synthetases of yeast and *E. coli*, with an observed 41% direct identity overall. Comparisons between the valyl- and isoleucyl-tRNA synthetases of *E. coli* yielded the highest level of homology detected between heterologous enzymes (19.2% direct identity overall). An alignment is presented between the three branched-chain aminoacyl-tRNA synthetases (valyl- and isoleucyl-tRNA synthetases of *E. coli* and yeast mitochondrial leucyl-tRNA synthetase) illustrating the close relatedness of these enzymes. These results give credence to the supposition that the branched-chain aminoacyl-tRNA synthetases along with methionyl-tRNA synthetase form a family of genes within the aminoacyl-tRNA synthetases that evolved from a common ancestral progenitor gene.

As a group the aminoacyl-tRNA synthetases of *Escherichia coli* are responsible for performing the same essential cellular function, the aminoacylation of tRNA. However, among the individual members of this class of enzymes there exists a high degree of diversity with regards to the overall sizes (1) and quarternary structures (2). While some individual size differences may possibly be ascribed to additional domains which serve functions other than those immediately required for the catalysis of tRNA aminoacylation, such as subunit-subunit interaction (3), autoregulatory functions (4, 5), or protein folding and t-RNA conformation constraint domains (6), the fact that only limited primary sequence homology is observed in pairwise comparisons of the amino acid sequences of these enzymes is seemingly inconsistent with synthetases sharing a common evolutionary origin or relatedness. In the vast majority of cases only pockets of limited primary sequence homology, occurring predominately within the amino-terminal half of these enzymes, have been discerned in these comparisons. With the singular exception of a comparison between the primary structure of yeast valyl-tRNA synthetase and isoleucyl-tRNA synthetase of *E. coli* no extended regions of primary sequence homologies between heterologous aminoacyl-tRNA synthetases have been observed to date (7).

We report here the entire nucleotide sequence of the *valS* gene encoding valyl-tRNA synthetase along with the corresponding deduced amino acid sequence. Homology comparisons between the deduced primary structure of valyl-tRNA synthetase and the primary structures of the other known aminoacyl-tRNA synthetases are described. Additional corroborative evidence of the substantial degree of relatedness which exists between the heterologous valyl-tRNA and isoleucyl-tRNA enzymes is presented in comparisons of these two enzymes from *E. coli* K12. Common sequence homologies with other branch-chained aminoacyl-tRNA synthetases strongly support the hypothesis that these enzymes evolved from a common progenitor gene.

EXPERIMENTAL PROCEDURES

Materials—All restriction endonucleases and other enzymes were purchased from Boehringer Mannheim Biochemicals or New England Biolabs, Inc. All radioactive compounds were obtained from Amersham Corp. All chemicals were either from Sigma or Mallinckrodt Chemical Works.

Nucleotide Sequencing Analysis—DNA sequence analysis was performed according to the dideoxy chain termination method of Sanger *et al.* (8). The identity of each nucleotide of the noncoding strand was verified by the independent determination of the complete DNA sequence of both strands, with some portions of each strand repeatedly analyzed from overlapping sequential deletions as illustrated in Fig. 1. The method of Henikoff (9) was employed for the generation of the M13 sequential deletion derivatives utilized in DNA sequence analysis. Screening of the potential sequential deletion derivatives prior to DNA sequence analysis was accomplished by either using representative cloned derivatives as template DNAs in dideoxy-T sequencing reactions (10), followed by electrophoresis in a buffer gradient gel (11) and autoradiography to allow banding pattern comparisons to be made, or by determining the relative sizes of the purified single-stranded deletion DNAs directly by electrophoresis in a 1.0% agarose gel buffered with 2 × Hellings (12).

DNA—M13 RF DNA (13) and cesium chloride band-purified plasmid DNA were prepared by standard methods (14).

Computer Analysis of Nucleotide and Amino Acid Sequences—Analyses of the determined nucleotide sequences were facilitated by use of the DNA Inspector II program (15). Analyses of the deduced amino acid sequence of valyl-tRNA synthetase and comparisons of the primary structure of valyl-tRNA synthetase with the other known aminoacyl-tRNA synthetase primary structures were accomplished by use of the programs (GENED, SEQ, and PEP) from BIONET® Intelligenetics (16) along with the programs (Codon Frequency, Best

* This work was supported in part by Grant GM 24330 from the National Institutes of Health. The programs GENED, SEQ, and PEP from Bionet® Intelligenetics were supported by National Institutes of Health Grant 1 041 RR-01685-05. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) J03497.

‡ Recipient of a CUPP graduate fellowship in biotechnology.

§ To whom correspondence should be addressed.

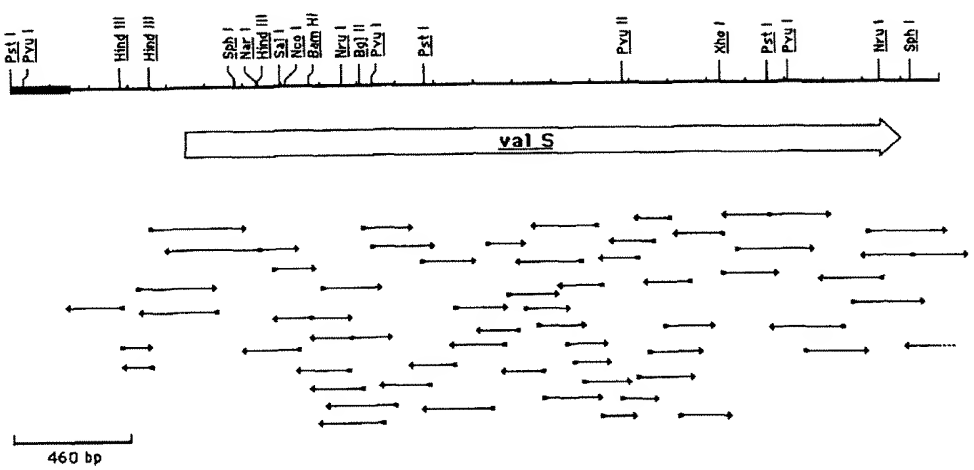


FIG. 1. **Partial restriction endonuclease map and DNA sequencing strategy of the *valS* gene of *E. coli* K12.** The *thicker lined* portion of the restriction map represents DNA sequences of the naturally occurring ColE1 plasmid portion present in the Clarke-Carbon library plasmid pLC26-22 (18). Recombinant M13mp10 and 11 or M13mp18 and 19 containing either discrete restriction fragments or sequential derivatives (9) of DNA excised from plasmid pLC26-22 were used as templates in DNA sequence determinations. The extent of the *arrows* beneath the restriction map represents the breadth of readable DNA sequence determined by the dideoxy chain termination method (8). The large *open arrow* below the partial restriction endonuclease map designates the DNA region encoding the *valS* structural gene.

Fit, Gap, and PepPlot) of the University of Wisconsin Genetics Computer Group (17).

RESULTS

Determination of the Nucleotide Sequence of the valS Gene of E. coli—Starting with a hybrid plasmid of the Clarke-Carbon *E. coli* library (18), the *valS* structural gene was subcloned and molecular genetic elements responsible for *valS* expression were characterized (19). Based on the physical map of the *valS* gene, specific DNA restriction endonuclease fragments were isolated and inserted into bacteriophage M13mp10 or mp11. The nucleotide sequences of these *valS* gene restriction endonuclease fragments were determined by the dideoxy-chain termination method (8). Additionally, larger-sized DNA restriction endonuclease fragments encompassing the distal four-fifths of the *valS* structural gene (ranging from 1.4 to 2.2 kb[1] in size) were isolated and inserted into bacteriophage M13mp18 or mp19. The replicative forms (RF) of these M13 bacterial phages were utilized to generate a series of sequential deletion derivatives spanning both strands of the DNA encoding valyl-tRNA synthetase (9). The nucleotide sequences of these *valS* gene M13 deletion derivatives were also determined by the method of Sanger *et al.* (8). The sequencing strategy employed illustrates that the nucleotide sequence was independently obtained for both strands of the DNA with much of the nucleotide sequence of each strand repeatedly determined from analysis of overlapping sequential DNA segments (Fig. 1). The 2856 nucleotide DNA sequence of the sense strand of the *valS* gene is shown in Fig. 2.

Localization and Characterization of the valS Gene—The deduced amino acid sequence of the *valS* gene is shown immediately below the determined nucleotide sequence in Fig. 2. Determination of the purified valyl-tRNA synthetase protein amino-terminal sequence[2] has confirmed the proposed translational start of *valS* (19). An open reading frame of 2856 nucleotides, extending from the amino-terminal methionine codon, encodes a deduced polypeptide of 951 amino acids in length. The calculated molecular weight, 108,070, is in close approximation to the previously determined molecular weight of 110,000 for valyl-tRNA synthetase (20). The deduced amino acid composition of the *valS* structural gene is in very close agreement with the amino acid composition obtained from protein hydrolysis of purified valyl-tRNA synthetase (the observed differences between the total percentages of individual amino acids is ≤1.1%).[2] It should be pointed out that while the deduced and hydrolyzed purified protein amino acid compositions are in close agreement they both differ markedly from the previously determined amino acid composition for valyl-tRNA synthetase of *E. coli* (20). A comparison of the codon frequency usage of *valS* with the average frequency of codon usage obtained from analysis of 25 abundant *E. coli* genes is presented in Table I. The percentage of codon usage for respective amino acids within *valS versus* the average utilization observed in 25 abundant genes shows the same general trends (21). Specifically, the frequency of rare codon usage in *valS* closely mimics the average observed in the other genes.

Finally, in contrast to earlier reports (22), the deduced *valS* primary sequence does not contain any significant repeat units. The purported existence of these repeat units, thought to be the result of a gene duplication/fusion event, was used to partially explain the large differences observed in the molecular weights within the aminoacyl-tRNA synthetases. While there is the hint of an internal repeat element within the *valS* deduced primary structure (at amino acid residues 328–343 and from 924 to 939, Fig. 2), the fact that there are no significant repeat units within this and other large aminoacyl-tRNA synthetases strongly argues that these polypeptides did not arise from a gene duplication/fusion event (23).

Primary Structure Homology between Valyl-tRNA Synthetase and Other Aminoacyl-tRNA Synthetases—Utilizing available computer programs (*cf.* "Materials and Methods"), we have compared the deduced primary structure of the *valS* gene with the primary structures of the *alaS*, *glnS*, *gltX*, *glyS*, *hisS*, *ileS*, *metG*, *pheS*, *serS*, *thrS*, *trpS*, and *tyrS* genes of *E. coli* and the *MSL1* and *VASI* genes of *Saccharomyces cerevisiae*, which encode yeast mitochondrial leucyl-tRNA synthetase and cytoplasmic valyl-tRNA synthetase, respectively (Refs. 4, 25–36, and 7, respectively). As expected, the strongest overall homology is detected when comparing the deduced primary structures of the unrelated homologous valyl-tRNA synthetase enzymes. Based on the sequence alignment shown in Fig. 3, there is a 41% overall direct amino acid correspondence between the two deduced primary sequences of valyl-tRNA synthetase obtained from yeast and bacteria. When the percent direct amino acid identity is calculated only for the amino proximal two-thirds of the two primary sequences the identity level rises to approximately 48.3%, reflecting the fact that the more strongly conserved regions are found toward the respective amino termini, the carboxyl-terminal portions

[1] The abbreviations used are: kb, kilobase; bp, base pair.
[2] W.-C. Chu and J. Horowitz, personal communication.

valS

```
        100                                 130                                 160
ATG GAA AAG ACA TAT AAC CCA CAA GAT ATC GAA CAG CCG CTT TAC GAG CAC TGG GAA AAG CAG GGC TAC TTT AAG CCT AAT GGC GAT GAA
Met Glu Lys Thr Tyr Asn Pro Gln Asp Ile Glu Gln Pro Leu Tyr Glu His Trp Glu Lys Gln Gly Tyr Phe Lys Pro Asn Gly Asp Glu
  1                                  10                                  20                                  30
        190                                 220                                 250
AGC CAG GAA AGT TTC TGC ATC ATG ATC CCG CCG CCG AAC GTC ACC GGC AGT TTG CAT ATG GGT CAC GCC TTC CAG CAA ACC ATC ATG GAT
Ser Gln Glu Ser Phe Cys Ile Met Ile Pro Pro Pro Asn Val Thr Gly Ser Leu His Met Gly His Ala Phe Gln Gln Thr Ile Met Asp
                                   40                                  50                                  60
        280                                 310                                 340
ACC ATG ATC CGC TAT CAG CGC ATG CAG GGC AAA AAC ACC CTG TGG CAG GTC GGT ACT GAC CAC GCC GGG ATC GCT ACC CAG ATG GTC GTT
Thr Met Ile Arg Tyr Gln Arg Met Gln Gly Lys Asn Thr Leu Trp Gln Val Gly Thr Asp His Ala Gly Ile Ala Thr Gln Met Val Val
                                   70                                  80                                  90
        370                                 400                                 430
GAG CGC AAG ATT GCC GCA GAA GAA GGT AAA ACC CGT CAC GAC TAC GGC GCC GAA GCT TTC ATC GAC AAA ATC TGG GAA TGG AAA GCG GAA
Glu Arg Lys Ile Ala Ala Glu Glu Gly Lys Thr Arg His Asp Tyr Gly Ala Glu Ala Phe Ile Asp Lys Ile Trp Glu Trp Lys Ala Glu
                                  100                                 110                                 120
        460                                 490                                 520
TCT GGC GGC ACC ATT ACC CGT CAG ATG CGC CGT CTC GGC AAC TCC GTC GAC TGG GAG CGT GAA CGC TTC ACC ATG GAC GAA GGC CTG TCC
Ser Gly Gly Thr Ile Thr Arg Gln Met Arg Arg Leu Gly Asn Ser Val Asp Trp Glu Arg Glu Arg Phe Thr Met Asp Glu Gly Leu Ser
                                  130                                 140                                 150
        550                                 580                                 610
AAT GCG GTG AAA GAA GTT TTC GTT CGT CTG TAT AAA GAA GAC CTG ATT TAC CGT GGC AAA CGC CTG GTA AAC TGG GAT CCG AAA CTG CGC
Asn Ala Val Lys Glu Val Phe Val Arg Leu Tyr Lys Glu Asp Leu Ile Tyr Arg Gly Lys Arg Leu Val Asn Trp Asp Pro Lys Leu Arg
                                  160                                 170                                 180
        640                                 670                                 700
ACC GCT ATC TCT GAC CTG GAA GTC GAA AAC CGC GAA TCG AAA GGT TCG ATG TGG CAC ATC CGC TAT CCG CTG GCT GAC GGT GCG AAA ACC
Thr Ala Ile Ser Asp Leu Glu Val Glu Asn Arg Glu Ser Lys Gly Ser Met Trp His Ile Arg Tyr Pro Leu Ala Asp Gly Ala Lys Thr
                                  190                                 200                                 210
        730                                 760                                 790
GCA GAC GGT AAA GAT TAT CTG GTG GTC GCG ACT ACC CGT CCA GAA ACC CTG CTG GGC GAT ACT GGC GTA GCC GTT AAC CCG GAA GAT CCG
Ala Asp Gly Lys Asp Tyr Leu Val Val Ala Thr Thr Arg Pro Glu Thr Leu Leu Gly Asp Thr Gly Val Ala Val Asn Pro Glu Asp Pro
                                  220                                 230                                 240
        820                                 850                                 880
CGT TAC AAA GAT CTG ATT GGC AAA TAT GTC ATT CTG CCG CTG GTT AAC CGT CGT ATT CCG ATC GTT GGC GAC GAA CAC GCC GAC ATG GAA
Arg Tyr Lys Asp Leu Ile Gly Lys Tyr Val Ile Leu Pro Leu Val Asn Arg Arg Ile Pro Ile Val Gly Asp Glu His Ala Asp Met Glu
                                  250                                 260                                 270
        910                                 940                                 970
AAA GGC ACC GGC TGC GTG AAA ATC ACT CCG GCG CAC GAC TTT AAC GAC TAT GAA GTG GGT AAA CGT CAG GCC CTG CCG ATG ATC AAC ATC
Lys Gly Thr Gly Cys Val Lys Ile Thr Pro Ala His Asp Phe Asn Asp Tyr Glu Val Gly Lys Arg His Ala Leu Pro Met Ile Asn Ile
                                  280                                 290                                 300
       1000                                1030                                1060
CTG ACC TTT GAC GGC GAT ATC CGT GAA AGC GCC CAG GTG TTC GAT ACC AAA GGT AAC GAA TCT GAC GTT TAT TCC AGC GAA ATC CCT GCA
Leu Thr Phe Asp Gly Asp Ile Arg Glu Ser Ala Gln Val Phe Asp Thr Lys Gly Asn Glu Ser Asp Val Tyr Ser Ser Glu Ile Pro Ala
                                  310                                 320                                 330
       1090                                1120                                1150
GAG TTC CAG AAA CTG GAG CGT TTT GCT GCA CGT AAA GCA GTC GTT GCC GCA GTT GAC GCG CTT GGC CTG CTG GAA GAA ATT AAA CCG CAC
Glu Phe Gln Lys Leu Glu Arg Phe Ala Ala Arg Lys Ala Val Val Ala Ala Val Asp Ala Leu Gly Leu Leu Glu Glu Ile Lys Pro His
                                  340                                 350                                 360
       1180                                1210                                1240
GAC CTG ACC GTT CCT TAC GGC GAC CGT GGC GGC GTA GTT ATC GAA CCA ATG CTG ACC GAC CAG TGG TAC GTG CGT GCC GAT GTC CTG GCG
Asp Leu Thr Val Pro Tyr Gly Asp Arg Gly Gly Val Val Ile Glu Pro Met Leu Thr Asp Gln Trp Tyr Val Arg Ala Asp Val Leu Ala
                                  370                                 380                                 390
       1270                                1300                                1330
AAA CCG GCG GTT GAA GCG GTT GAG AAC GGC GAC ATT CAG TTC GTA CCG AAG CAG TAC GAA AAC ATG TAC TTC TCC TGG ATG CGC GAT ATT
Lys Pro Ala Val Glu Ala Val Glu Asn Gly Asp Ile Gln Phe Val Pro Lys Gln Tyr Glu Asn Met Tyr Phe Ser Trp Met Arg Asp Ile
                                  400                                 410                                 420
       1360                                1390                                1420
CAG GAC TGG TGT ATC TCT CGT CAG TTG TGG TGG GGT CAC CGT ATC CCG GCA TGG TAT GAC GAA GCG GGT AAC GTT TAT GTT GGC CGC AAC
Gln Asp Trp Cys Ile Ser Arg Gln Leu Trp Trp Gly His Arg Ile Pro Ala Trp Tyr Asp Glu Ala Gly Asn Val Tyr Val Gly Arg Asn
                                  430                                 440                                 450
       1450                                1480                                1510
GAA GAC GAA GTG CGT AAA GAA AAT AAC CTC GGT GCT GAT GTT GTC CTG CGT CAG GAC GAA GAC GTT CTC GAT ACC TGG TTC TCT TCT GCG
Glu Asp Glu Val Arg Lys Glu Asn Asn Leu Gly Ala Asp Val Val Leu Arg Gln Asp Glu Asp Val Leu Asp Thr Trp Phe Ser Ser Ala
                                  460                                 470                                 480
       1540                                1570                                1600
CTG TGG ACC TTC TCT ACC CTT GGC TGG CCG GAA AAT ACC GAC GCC CTG CGT CAG TTC CAC CCA ACC AGC GTG ATG GTA TCT GGT TTC GAC
Leu Trp Thr Phe Ser Thr Leu Gly Trp Pro Glu Asn Thr Asp Ala Leu Arg Gln Phe His Pro Thr Ser Val Met Val Ser Gly Phe Asp
                                  490                                 500                                 510
       1630                                1660                                1690
ATC ATT TTC TTC TGG ATT GCC CGC ATG ATC ATG ATG ACC ATG CAC TTC ATC AAA GAT GAA AAT GGC AAA CCG CAG GTG CCG TTC CAC ACC
Ile Ile Phe Phe Trp Ile Ala Arg Met Ile Met Met Thr Met His Phe Ile Lys Asp Glu Asn Gly Lys Pro Gln Val Pro Phe His Thr
                                  520                                 530                                 540
       1720                                1750                                1780
GTT TAC ATG ACC GGC CTG ATT CGT GAT GAC GAA GGC CAG AAG ATG TCC AAA TCC AAG GGT AAC GTT ATC GAC CCA CTG GAT ATG GTT GAC
Val Tyr Met Thr Gly Leu Ile Arg Asp Asp Glu Gly Gln Lys Met Ser Lys Ser Lys Gly Asn Val Ile Asp Pro Leu Asp Met Val Asp
                                  550                                 560                                 570
       1810                                1840                                1870
GGT ATT TCG CTG CCA GAA CTG CTG GAA AAA CGT ACC GGC AAT ATG ATG CAG CCG CAG CTG GCG GAC AAA ATC CGT AAG CGC ACC GAG AAG
Gly Ile Ser Leu Pro Glu Leu Leu Glu Lys Arg Thr Gly Asn Met Met Gln Pro Gln Leu Ala Asp Lys Ile Arg Lys Arg Thr Glu Lys
                                  580                                 590                                 600
       1900                                1930                                1960
CAG TTC CCG AAC GGT ATT GAG CCG CAC GGT ACT GAC GCG CTG CGC TTC ACC CTG GCG GCG CTG GCG TCT ACC GGT CGT GAC ATC AAC TGG
Gln Phe Pro Asn Gly Ile Glu Pro His Gly Thr Asp Ala Leu Arg Phe Thr Leu Ala Ala Leu Ala Ser Thr Gly Arg Asp Ile Asn Trp
                                  610                                 620                                 630
       1990                                2020                                2050
GAT ATG AAG CGT CTG GAA GGT TAC CGT AAC TTC TGT AAC AAG CTG TGG AAC GCC AGC CGC TTT GTC CTG ATG AAC ACA GAA GGT CAG GAT
Asp Met Lys Arg Leu Glu Gly Tyr Arg Asn Phe Cys Asn Lys Leu Trp Asn Ala Ser Arg Phe Val Leu Met Asn Thr Glu Gly Gln Asp
                                  640                                 650                                 660
       2080                                2110                                2140
TGC GGC TTC AAC GGC GGC GAA ATG ACG CTG TCG CTG GCG GAC CGC TGG ATT CTG GCG GAG TTC AAC CAG ACC ATC AAA GCG TAC CGC GAA
Cys Gly Phe Asn Gly Gly Glu Met Thr Leu Ser Leu Ala Asp Arg Trp Ile Leu Ala Glu Phe Asn Gln Thr Ile Lys Ala Tyr Arg Glu
                                  670                                 680                                 690
       2170                                2200                                2230
GCG CTG GAC AGC TTC CGC TTC GAT ATC GCC GCA GGC ATT CTG TAT GAG TTC ACC TGG AAC CAG TTC TGT GAC TGG TAT CTC GAG CTG ACC
Ala Leu Asp Ser Phe Arg Phe Asp Ile Ala Ala Gly Ile Leu Tyr Glu Phe Thr Trp Asn Gln Phe Cys Asp Trp Tyr Leu Glu Leu Thr
                                  700                                 710                                 720
```

FIG. 2. **The nucleotide sequence and deduced amino acid sequence of the nontranscribed DNA strand of the *valS* gene of *E. coli* K12.** The complete nucleotide sequence of the *valS* gene, beginning with the translational start codon (ATG) at nucleotide position +93, is listed. The nucleotide numbering is relative to the start of transcription (19). The predicted amino acid sequence of valyl-tRNA synthetase, the *valS* gene product, is listed immediately below the nucleotide sequence with the residues numbered from the start of translation. The sequence of the first deduced 10 amino acids agrees with the determined amino-terminal sequence of the enzyme.[2]

```
      2260                                    2290                                    2320
AAG CCG GTA ATG AAC GGT GGC ACC GAA GCA GAA CTG CGC GGT ACT CGC CAT ACG CTG GTG ACT GTA CTG GAA GGT CTG CTG CGC CTC GCG
Lys Pro Val Met Asn Gly Gly Thr Glu Ala Glu Leu Arg Gly Thr Arg His Thr Leu Val Thr Val Leu Glu Gly Leu Leu Arg Leu Ala
                                730                                     740                                     750
      2350                                    2380                                    2410
CAT CCG ATC ATT CCG TTC ATC ACC GAA ACC ATC TGG CAG CGT GTG AAA GTA CTT TGC GGT ATC ACT GCC GAC ACC ATC ATG CTG CAG CCG
His Pro Ile Ile Pro Phe Ile Thr Glu Thr Ile Trp Gln Arg Val Lys Val Leu Cys Gly Ile Thr Ala Asp Thr Ile Met Leu Gln Pro
                                760                                     770                                     780
      2440                                    2470                                    2500
TTC CCG CAG TAC GAT GCA TCT CAG GTT GAT GAA GCC GCA CTG GCC GAC ACC GAA TGG CTG AAA CAG GCG ATC GTT GCG GTA CGT AAC ATC
Phe Pro Gln Tyr Asp Ala Ser Gln Val Asp Glu Ala Ala Leu Ala Asp Thr Glu Trp Leu Lys Gln Ala Ile Val Ala Val Arg Asn Ile
                                790                                     800                                     810
      2530                                    2560                                    2590
CGT GCA GAA ATG AAC ATC GCG CCG GGC AAA CCG CTG GAG CTG CTG CTG CGT GGT TGC AGC GCG GAT CGA GAA CGT CGC GTA AAT GAA AAC
Arg Ala Glu Met Asn Ile Ala Pro Gly Lys Pro Leu Glu Leu Leu Leu Arg Gly Cys Ser Ala Asp Arg Glu Arg Arg Val Asn Glu Asn
                                820                                     830                                     840
      2620                                    2650                                    2680
CGT GGC TTC CTG CAA ACC CTG GCG CGT CTG GAA AGT ATC ACC GTG CTG CCT GCC GAT GAC AAA GGT CCG GTT TCC GTT ACG AAG ATC ATC
Arg Gly Phe Leu Gln Thr Leu Ala Arg Leu Glu Ser Ile Thr Val Leu Pro Ala Asp Asp Lys Gly Pro Val Ser Val Thr Lys Ile Ile
                                850                                     860                                     870
      2710                                    2740                                    2770
GAC GGT GCA GAG CTG CTG ATC CCG ATG GCT GGC CTC ATC AAC AAA GAA GAT GAG CTG GCG CGT CTG GCG AAA GAA GTG GCG AAG ATT GAA
Asp Gly Ala Glu Leu Leu Ile Pro Met Ala Gly Leu Ile Asn Lys Glu Asp Glu Leu Ala Arg Leu Ala Lys Glu Val Ala Lys Ile Glu
                                880                                     890                                     900
      2800                                    2830                                    2860
GGT GAA ATC AGC CGT ATC GAG AAC AAA CTG GCG AAC GAA GGC TTT GTC GCC CGC GCA CCG GAA GCG GTC ATC GCG AAA GAG CGT GAG AAG
Gly Glu Ile Ser Arg Ile Glu Asn Lys Leu Ala Asn Glu Gly Phe Val Ala Arg Ala Pro Glu Ala Val Ile Ala Lys Glu Arg Glu Lys
                                910                                     920                                     930
      2890                                    2920                                    2950
CTG GAA GGC TAT GCG GAA GCG AAA GCG AAA CTG ATT GAA CAG CAG GCT GTT ATC GCC GCG CTG TAA
Leu Glu Gly Tyr Ala Glu Ala Lys Ala Lys Leu Ile Glu Gln Gln Ala Val Ile Ala Ala Leu  *
                                940                                     950
```

FIG. 2—*continued*

TABLE I

Codon frequency usage of the valS gene as compared with codon usage in 25 E. coli genes

E. coli codon usage compilation from Konigsberg and Godson (21).

Residue and codon	Codon frequencies[a]		Residue and codon	Codon frequencies[a]	
	ValS	*E. coli*		*ValS*	*E. coli*
Phe UUU	6 (17)	104 (44)	Tyr UAU	13 (50)	69 (41)
Phe UUC	29 (83)	135 (56)	Tyr UAC	13 (50)	101 (59)
Leu UUA	0 (0)	36 (6)	Ter UAA	1	22
Leu UUG	2 (3)	51 (8)	Ter UAG	0	1
Leu CUU	4 (5)	54 (9)	Ter UGA	0	2
Leu CUC	6 (8)	41 (7)			
Leu CUA	0 (0)	11 (2)	His CAU	3 (18)	42 (39)
Leu CUG	67 (85)	432 (69)	His CAC	14 (82)	66 (61)
Ile AUU	19 (30)	151 (37)	Gln CAA	3 (8)	75 (27)
Ile AUC	44 (70)	252 (62)	Gln CAG	33 (92)	207 (73)
Ile AUA	0 (0)	2 (1)			
			Asn AAU	7 (18)	57 (24)
Met AUG	32	189	Asn AAC	33 (83)	179 (76)
Val GUU	25 (41)	182 (38)	Lys AAA	38 (72)	296 (77)
Val GUC	12 (20)	62 (13)	Lys AAG	15 (28)	90 (23)
Val GUA	10 (16)	111 (23)			
Val GUG	14 (23)	130 (27)	Asp GAU	25 (38)	175 (51)
			Asp GAC	40 (62)	168 (49)
Ser UCU	10 (31)	86 (27)			
Ser UCC	7 (22)	83 (26)	Glu GAA	63 (79)	328 (73)
Ser UCA	0 (0)	27 (8)	Glu GAG	17 (21)	119 (27)
Ser UCG	4 (13)	37 (11)			
Ser AGU	3 (9)	21 (6)	Cys UGU	3 (38)	21 (42)
Ser AGC	8 (25)	70 (22)	Cys UGC	5 (63)	29 (58)
Pro CCU	4 (10)	24 (9)	Trp UGG	24	48
Pro CCC	0 (0)	16 (6)			
Pro CCA	6 (14)	53 (20)	Arg CGU	37 (61)	201 (58)
Pro CCG	32 (76)	174 (65)	Arg CGC	23 (38)	121 (35)
			Arg CGA	1 (2)	8 (2)
Thr ACU	8 (15)	76 (24)	Arg CGG	0 (0)	11 (3)
Thr ACC	40 (75)	162 (51)	Arg AGA	0 (0)	4 (1)
Thr ACA	2 (4)	19 (6)	Arg AGG	0 (0)	1 (0.25)
Thr ACG	3 (6)	63 (20)			
			Gly GGU	27 (41)	231 (48)
Ala GCU	8 (10)	202 (28)	Gly GGC	38 (58)	197 (41)
Ala GCC	20 (26)	136 (19)	Gly GGA	0 (0)	22 (5)
Ala GCA	14 (18)	166 (23)	Gly GGG	1 (2)	33 (7)
Ala GCG	36 (46)	221 (30)			
			Total	951[b]	6,478[b]

[a] Numbers represent times codon used in genes. Numbers in parentheses represent the percentage of codon usage for the respective amino acid within *valS* or the 25 grouped *E. coli* genes.

[b] Total codons minus translational stop codon(s).

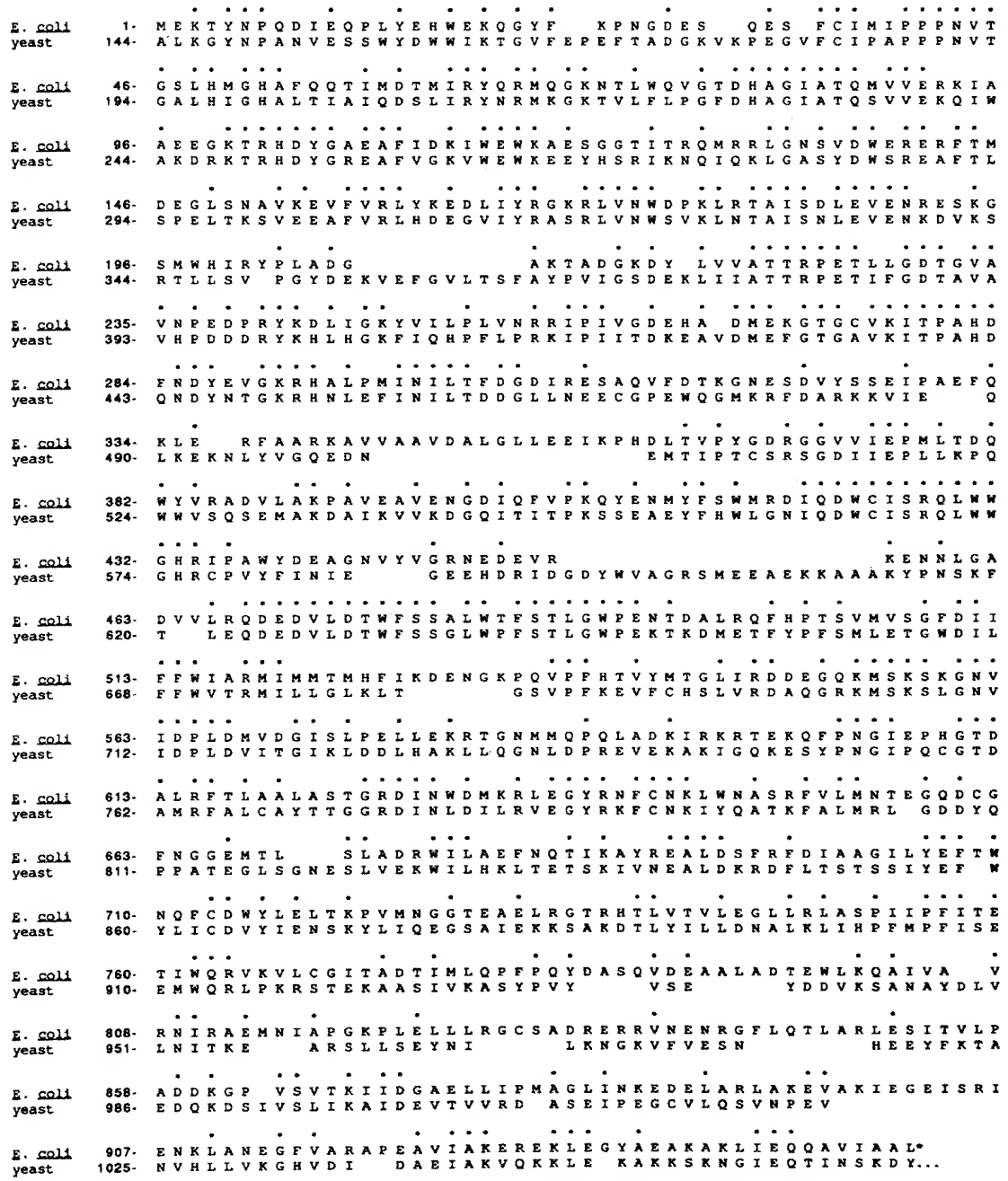


FIG. 3. **Primary sequence homology alignment of valyl-tRNA synthetase of *E. coli* and the yeast cytoplasmic valyl-tRNA synthetase of *S. cerevisiae*.** The deduced primary sequences of the *valS* gene of *E. coli* and the *VAS1* gene of *S. cerevisiae*, both encoding valyl-tRNA synthetase, are aligned (7). Identical amino acids are indicated by a *filled circle* located immediately above identical residues within the two aligned sequences. The numbers to the left of each line give the residue position number of the first amino acid listed relative to the start of each respective primary sequence. With regard to previously identified or proposed functionally equivalent catalytic and/or binding residues, there is a 12/14 match found at the consensus HIGH region (*E. coli* sequence positions 40–53), a 16/17 match is found at the DWCISRQL consensus sequence (*E. coli* sequence positions 420–436) and a 14/16 match at the proposed KMSK consensus site thought responsible (40) for binding the 3′-end of the tRNA (*E. coli* sequence position 552–567). An additional region of sustantial homology has a 22/24 direct correspondence (*E. coli* sequence positions (468–491). The carboxyl terminus of *E. coli* primary sequence is indicated by an *asterisk*.

apparently having diverged more throughout evolution. It is of interest to note that the yeast valyl-tRNA synthetase enzyme has additional sequence elements located at both termini that apparently are not found within the equivalent bacterial enzyme, approximately 140 residues at the amino-terminal and 40 residues at the carboxyl-terminal (Fig. 3). Whether or not these addition sequences in yeast are due merely to species differentiation or might serve some additional functional role is conjecture at this time.

As previously mentioned, the strongest overall primary

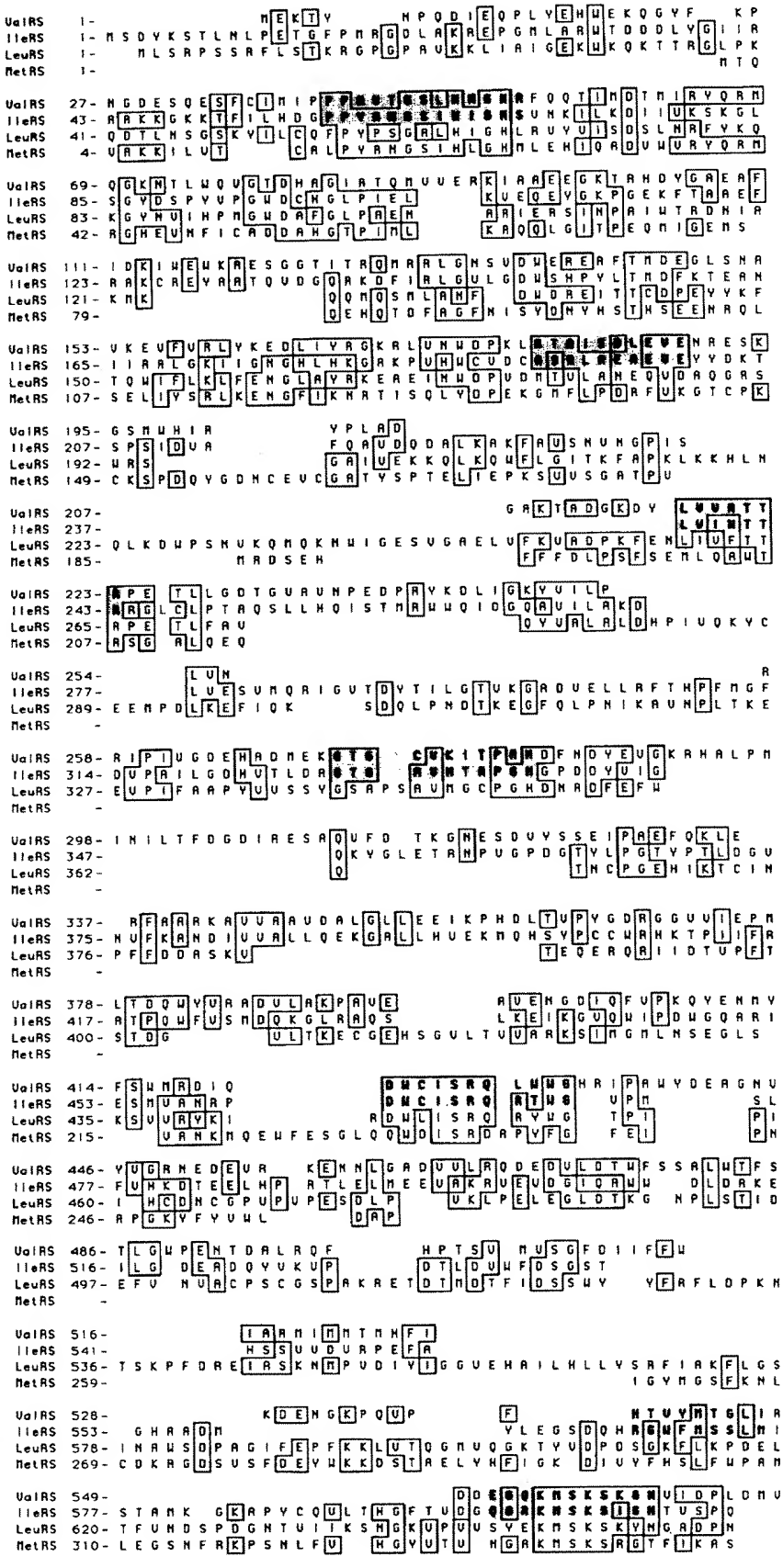


FIG. 4. **Primary sequence homology alignment of the valyl-, isoleucyl-, and methionyl-tRNA synthetases of *E. coli* and the yeast mitochondrial leucyl-tRNA synthetase of *S. cerevisiae*.** The deduced primary sequence of the *valS* gene encoding valyl-tRNA synthetase (*ValRS*) is aligned with the primary sequences of isoleucyl- (*IleRS*) and methionyl-tRNA synthetases (*MetRS*) of *E. coli* along with the primary sequence of the yeast mitochondrial leucyl-tRNA synthetase (*LeuRS*) (Refs. 29, 30 and 36, respectively). Identical amino acid residues are *boxed* if two or more sequences have a common residue at the same alignment position. Twelve regions of substantial homology which exist between valyl-tRNA synthetase and isoleucyl-tRNA synthetase are indicated by *stippling* in both sequences. The carboxyl terminus of each respective primary sequence is indicated by an *asterisk*.

FIG. 4—*continued*

sequence homology observed in pairwise comparisons between heterologous aminoacyl-tRNA synthetases is detected when comparing the valyl-tRNA synthetase and isoleucyl-tRNA synthetase primary sequences of *E. coli*. Sequence homology comparisons between the valyl- and isoleucyl-tRNA synthetase enzymes, utilizing the depicted optimum valyl-, isoleucyl-, methionyl-tRNA synthetase/yeast mitochondrial leucyl-tRNA synthetase alignment (Fig. 4), show an overall 19.2% direct amino acid identity per unit length and a 41.0% "chemical equivalent" amino acid identity per unit length. While gapping was introduced to allow for insertions or deletions present within the four individual synthetases, these values are in close agreement with values obtained from the optimal alignment found when only the valyl-tRNA synthetase and isoleucyl-tRNA synthetase enzymes were compared (20% direct identity and 43.2% similar identity; alignment not shown). Both the comprehensive alignment of the four branched-chain aminoacyl-tRNA synthetases and the pairwise comparison between valyl- and isoleucyl-tRNA synthetase were substantially based on the alignment of 12 short blocks of relatively strong homology (>80% at the level of "chemically equivalent" amino acids; regions *stippled* in Fig. 4) that exist primarily between the closely related valyl- and isoleucyl-tRNA synthetase enzymes but are in many cases common to both, or one of, the remaining two synthetases, methionyl-tRNA synthetase and yeast mitochondrial leucyl-tRNA synthetase. The preservation of all, or a subset thereof, of these 12 conserved regions within the valyl-, methionyl-, isoleucyl-, and leucyl-tRNA synthetases strongly suggests that these functionally related enzymes may also form an evolutionarily related family within the aminoacyl-tRNA synthetases by virtue of having diverged from a common ancestral gene. The fact that these 12 homologous pockets still remain within this group of most ancient of proteins suggests that these regions possibly represent functionally related blocks within this family of aminoacyl-tRNA synthetases.

Significantly, these 12 sequentially ordered regions of substantial homology are spread out along the entire lengths of both valyl- and isoleucyl-tRNA synthetases (Fig. 4). This finding contrasts sharply with the results of all but one (7) of a number of previous alignment studies, where the limited amount of primary sequence homology observed in pairwise comparisons between heterologous aminoacyl-tRNA synthetases, if observed, was found to occur predominately within the amino proximal halves of these enzymes where the catalytic core domain is believed to reside (40).

As previously mentioned, a subset of the same 12 regions found common to both valyl-tRNA synthetase and isoleucyl-tRNA synthetase are also regions of substantial shared homology that are observed in a pairwise comparison between the deduced primary structures of valyl-tRNA synthetase and methionyl-tRNA synthetase of *E. coli*. However, in this case the six regions of substantial homology occur primarily within the amino-terminal half of the 677-amino acid long methionyl-tRNA synthetase. Several of the 12 sequentially ordered blocks found common to both valyl-tRNA synthetase and isoleucyl-tRNA synthetase do not appear to have substantially homologous regions within the methionyl-tRNA synthetase primary sequence (*i.e.* regions 2 and 4, valyl-tRNA synthetase residue numbers 180–189 and 272–282; Fig. 4). As

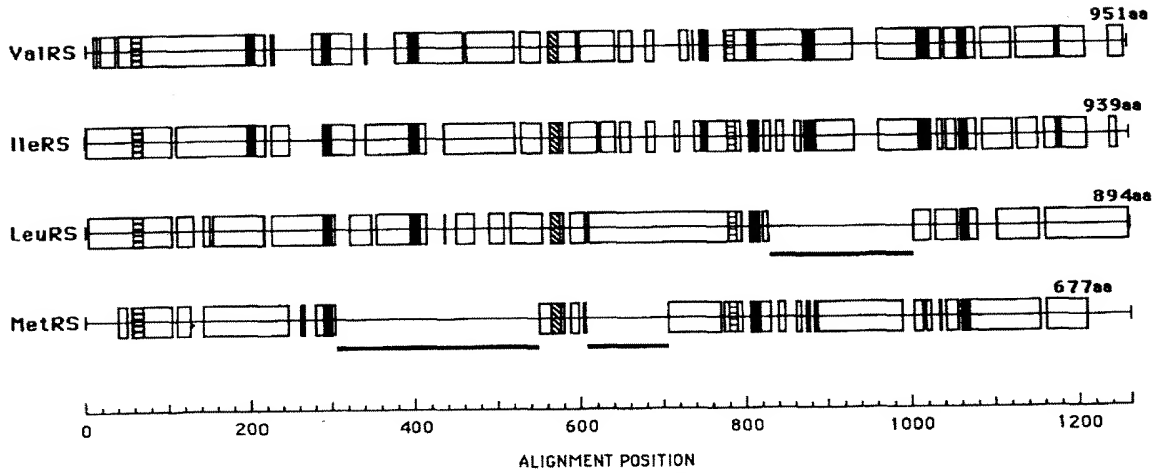


FIG. 5. **Schematic representation of the primary sequence alignment between valyl-, isoleucyl-, and methionyl-tRNA synthetases of *E. coli* and mitochondrial leucyl-tRNA synthetase of *S. cerevisiae* as depicted in Fig. 4.** A schematic representation is presented to illustrate homologous regions and gapping that was introduced to achieve the alignment between valyl-, isoleucyl-, and methionyl-tRNA synthetases of *E. coli* and mitochondrial leucyl-tRNA synthetase of *S. cerevisiae* presented in Fig. 4. The *rectangles* along each line represent scaled contiguous sequence elements that exist along the primary structure of each of the four aligned proteins. The *connecting line* between the rectangles indicates the extent of the gaps introduced to bring homologous regions into alignment. Indicated by either specific geometric patterns (*thick horizontal bars*, HIGH; *slanted hatched* lines, DWCISRQ, and *thin horizontal lines*, KMSK) or *solid black* within these rectangles are positions along each sequence that share sequence homology with the 12 substantial homologous regions which exist between valyl- and isoleucyl-tRNA synthetase (Fig. 4). The *thick bars* highlight regions missing in the sequence immediately above. *Scale* below gives overall position of the homologous blocks relative to the alignment of Fig. 4. The overall lengths of the individual synthetases are given immediately above and on the right.

illustrated schematically in Fig. 5, the latter of these two mentioned regions falls within an area common to the primary sequences of the other three synthetases but not present within the methionyl-tRNA synthetase primary sequence. These approximately 220-residue long inserts common to valyl-, isoleucyl-, and leucyl-tRNA synthetase appear to be the result of an extended connective polypeptide region not present within the region of the methionyl-tRNA synthetase primary sequence that forms the mononucleotide-binding fold believed to be involved in binding methionyl adenylate (41). The methionyl-tRNA synthetase primary sequence additionally lacks a smaller region of approximately 100 residues located close to the block (DWCISRQ) common to all four synthetases that immediately follows the previously defined missing region (Fig. 5). The overall sequence homology observed between valyl-tRNA synthetase and methionyl-tRNA synthetase, utilizing the alignment indicated in Fig. 4, is 20.2% similar and 9.5% direct identity.

In comparisons between the unrelated heterologous yeast mitochondrial (*S. cerevisiae*) leucyl-tRNA synthetase (39) and valyl-tRNA synthetase of *E. coli*, it is observed that a number of the previously defined blocks of shared homology, which exist between valyl-tRNA synthetase and isoleucyl-tRNA synthetase are also common to the yeast leucyl-tRNA synthetase (Fig. 4). Furthermore, the yeast leucyl-tRNA synthetase enzyme possesses regions that are homologous to the extended connective polypeptide region present in valyl-tRNA synthetase and isoleucyl-tRNA synthetase enzymes but not found as previously mentioned within the methionyl-tRNA synthetase enzyme. However, as illustrated in Fig. 5, there is a span of approximately 170 residues in length that is lacking in the primary sequence of the yeast sequence but present to greater or lesser extents within the bacterial primary sequences. The values obtained in pairwise comparisons between the primary structures of valyl- and leucyl-tRNA synthetase, based on the alignment of Fig. 4, are 27.1% similar and 14.6% direct identity.

Taken together, the observed degree of relatedness between these four enzymes lends credence to the supposition initially advanced by Wetzel (39) that valyl-, isoleucyl-, leucyl-, and methionyl-tRNA synthetase are all members of a family within the aminoacyl-tRNA synthetases. Based on the sequence homologies that exist within this proposed family the following evolutionary linkages are consistent: methionyl-tRNA synthetase → leucyl-tRNA synthetase → isoleucyl-tRNA synthetase and valyl-tRNA synthetase (Fig. 7).

With the exceptions of the previously mentioned isoleucyl- and methionyl-tRNA synthetase of *E coli* and the leucyl- and valyl-tRNA synthetase of yeast, comparisons between the primary structure of bacterial valyl-tRNA synthetase and the primary sequences encoded by the other listed genes (results not shown) detected only limited homology of a much lesser degree than was detected in comparisons with methionyl-tRNA synthetase. While no extensive segments of overall homology exist, there are several short blocks of equivalent homology that valyl-tRNA synthetase has in common with a number of the other synthetases. The more substantially shared regions of chemical equivalent amino acid homology which exist between these heterologous enzymes are illustrated in Fig. 6. The sequence homologous to the consensus or identity sequence HIGH, which is located near to the the amino-terminal end in all the aminoacyl-tRNA synthetases possessing this region and believed to play a role in the binding of ATP or the adenyl part of the adenylate intermediate (43), is also found in the valyl-tRNA synthetase primary sequence (Fig. 5). As illustrated in Fig. 6, the valyl-tRNA synthetase primary sequence differs in the most commonly variable residue of the consensus sequence with the chemically conserved substitution of a methionine residue for an isoleucine residue of the HIGH consensus sequence. The valyl-tRNA synthetase primary sequence also possesses a region of homology shared with many other synthetases which is believed to be involved with binding the 3'-end of the tRNA molecule (44). As illustrated in Figs. 4 and 5, this consensus sequence KMSKS is

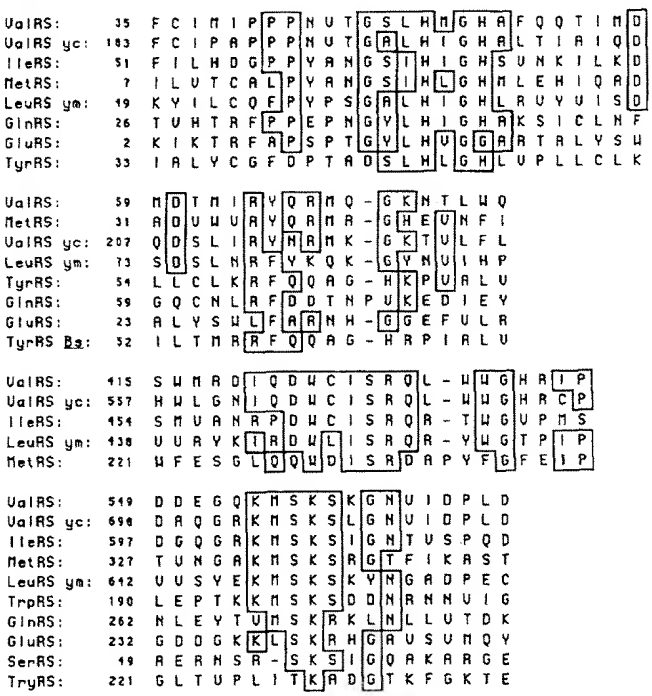


FIG. 6. **Amino acid sequence homology between valyl-tRNA synthetase (*ValRS*) and other known aminoacyl-tRNA synthetase primary sequence regions.** The corresponding aligned primary sequence regions show relatively strong identical and chemically equivalent amino acid homology with valyl-tRNA synthetase when allowing for minimal gapping. The *number* to the left of each sequence refers to the first listed residue position relative to the start of translation for each respective synthetase (Refs. listed in text). *Dashes* indicate gaps in sequence to maximize alignments. All synthetase sequences are *E. coli* derived unless otherwise noted. Abbreviations: *GlnRS*, glutaminyl-; *GluRS*, glutamyl-; *IleRS*, isoleucyl; *LeuRS*, leucyl-; *MetRS*, methionyl-; *SerRS*, seryl-; *TrpRS*, tryptophanyl-; and *TyrRS*, tyrosyl-; (*Bs*, *B. stearothermophilus; ym*, yeast mitochondrial, or *yc*, yeast cytoplasmic (*S. cerevisiae*)).

found in all four members of the proposed family, even though the enzymes are of prokaryotic and eukaryotic origins. Another region of previously unreported homology, which is common to all except isoleucyl-tRNA synthetase of the proposed family along with several other aminoacyl-tRNA synthetases, is also indicated in Fig. 4. Finally, there appears to be a region of sequence homology that is common only to the proposed family of branched-chain aminoacyl-tRNA synthetases. As illustrated in Fig. 4, this consensus sequence DWCISRQ, which in the case of isoleucyl-tRNA synthetase has been shown to possess an *N*-ethylmaleimide reactive cysteine that preferentially inactivates the isoleucyl-tRNA synthetase enzyme, is exactly identical to the homologous region present within in the valyl-tRNA synthetase enzyme that possess a 9 out of 11 residues direct identity with the isoleucyl-tRNA synthetase enzyme sequence (32).

DISCUSSION

In general, previous pairwise comparisons directed at uncovering possible sequence similarities between unrelated heterologous aminoacyl-tRNA synthetases, in the hope of defining domains involved in the binding of substrates and/or catalysis, have not revealed any extended regions of similarity; however, several synthetase pairs showed a number of short regions (6–14 residues) of statistically significant similarities (41). This is not surprising considering the fact that these

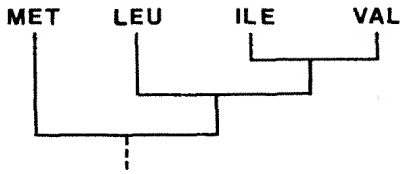


FIG. 7. **Depiction of the evolutionary relatedness occurring between the valyl-, isoleucyl-, and methionyl-tRNA synthetases of *E. coli* and yeast mitochondrial leucyl-tRNA synthetase.** Based on the percent sequence homology values obtained from analyses of the alignment depicted in Fig. 4, the degree of "evolutionary relatedness" occurring between the four compared aminoacyl-tRNA synthetases is presented. The drawing depicts an "unrooted" tree (cladogram) with branch lengths drawn proportionally on the *vertical axis* to depict the evolution of these related sequences, isoleucyl-tRNA synthetase (*IleRS*) and valyl- (*ValRS*), leucyl- (*LeuRS*), and methionyl-tRNA synthetases (*MetRS*) (46).

enzymes are representative of some of the most ancient of all proteins, which implies that these enzymes have experienced extensive multiple evolutionary replacements. Imposed on top of this background, the existence of additional domains that have been implicated in functions other than catalysis, such as subunit interaction or specific regulatory functions, means that only a portion of some of the primary structures of these enzymes can be reasonably expected to exhibit extensive homology (37). With these two caveats in mind it was somewhat unexpected to find the degree of chemically equivalent homology that exists between the branched-chain aminoacyl-tRNA synthetases. The observed 19.2% direct identity per unit length which exists between *E. coli* valyl- and isoleucyl-tRNA synthetases initially appears to be only moderately significant. However, when the length of the two primary structures are taken into account the authenticity of the relationship between valyl-tRNA synthetase and isoleucyl-tRNA synthetase is highly significant (42). In fact, when the average alignment score was computed from sets of scrambled sequences (whose compositions and lengths were both identical to valyl-tRNA synthetase and isoleucyl-tRNA synthetase and subjected to the same alignment and scoring procedure used for the authentic valyl/isoleucyl-tRNA synthetase alignment), it was found that the alignment score obtained for the genuine sequences was more than 9.0 standard deviations above the scrambled comparison average.[3] Moreover, when chemically equivalent amino acids are scored with the alignment shown in Fig. 4 there is a 41.0% direct correspondence between the primary sequences of valyl- and isoleucyl-tRNA synthetase. Clearly, the observed similarity between these two aminoacyl-tRNA synthetases is not due to chance but rather represents a genuine common ancestry. Additionally, the degree of relatedness is quite high among all individual pairings of the four synthetases (see Fig. 7) that comprise this proposed evolutionarily related family of synthetases (39). The fact that many of the ordered regions of substantial homology depicted in Fig. 4 are common to all four branched-chain aminoacyl-tRNA synthetases (Fig. 5) indicates these synthetases are more closely related than other known synthetase groupings and that these remaining pockets of sequence similarity possibly represent functionally important segments contributing to the function of these heterologous enzymes.

As previously mentioned and illustrated in the schematic of Fig. 5, the valyl-, isoleucyl-, and leucyl-tRNA synthetase enzymes have an additional peptide loop separating domains that are equivalent to regions of the methionyl-tRNA synthe-

[3] R. F. Doolittle, personal communication.

tase enzyme believed to be involved in the formation of the adenylate-binding fold (38). It would be of interest to see if presence of these extented connnective polypeptide regions is required for formation of the adenylate or the subsequent aminoacylation of tRNA. Isolation of genetically engineered deletion mutants spanning this region should provide information about the structural function of this extented connective polypeptide domain.

It should be noted that while short regions of significant homology exist within the carboxyl thirds of these enzymes, specifically between valyl- and isoleucyl-tRNA synthetases (Fig. 3), the fact remains that more extensively shared homology is present within the amino proximal thirds of these enzymes. This is taken to mean that within these amino proximal regions, which are quite homologous to the methionyl-tRNA synthetase and thus comparable to the *Bacillus stearothermophilus* tyrosyl-tRNA synthetase by x-ray diffraction studies (45), that the tertiary structures of these four synthetases should be quite homologous.

A systematic search for amino acid sequences that potentially could form metal-binding domains in nucleic acid-binding proteins has identified such proposed sequences in several of the aminoacyl-tRNA synthetases, specifically both methionyl- and isoleucyl-tRNA synthetases of *E. coli* have been so identified (43). These sequences, of the form Cys-X_2-Cys-X_{9-16}-Cys-X_2-Cys are thought to bind the one Zn^{2+} ion which is found per polypeptide chain in both methionyl- and isoleucyl-tRNA synthetase proteins of *E. coli* (28, 29). A search of the deduced valyl-tRNA synthetase primary structure for sequences with 4 Cys or His residues arranged in a manner suggestive of a metal-binding domain found no corresponding sequences present within valyl-tRNA synthetase. This finding is in contrast to reports that all three thermostable valyl-, isoleucyl-, and methionyl-tRNA synthetases of *Thermus thermophilus* HB8 bind Zn^{2+} ions (44). The proposed metal-binding domains of both isoleucyl- and methionyl-tRNA synthetase are located in distinctly different regions of their respective sequences. The proposed domain of the isoleucyl-tRNA synthetase enzyme is located proximal to the carboxyl terminus of the enzyme (isoleucyl-tRNA synthetase, residues 902–925), while the proposed methionyl-tRNA synthetase metal-binding domain is located within the amino-terminal third of the enzyme (methionyl-tRNA synthetase, residues 145–161). Therefore, it seems that if indeed functional, the proposed metal-binding domains found within these two enzymes are representative of a more recent evolutionary acquisition since 1) these sequences are not found within the primary sequences of other closely related family members (*i.e. E. coli* valyl-tRNA synthetase and yeast mitochondrial leucyl-tRNA synthetase and cytoplasmic valyl-tRNA synthetase and more importantly, 2) these proposed sequences are positioned in an order different than other ordered regions of shared sequence homology. This suggests that these proposed metal-binding domains were not present in the ancestral progenitor gene responsible for the methionyl-, leucyl-, valyl-, and isoleucyl-tRNA synthetase family.

Acknowledgments—We are indebted to Russel F. Doolittle for useful discussions and advice concerning the initial full-length primary sequence alignment of valyl-tRNA synthetase and isoleucyl-tRNA synthetase. Additionally, we thank Alexander Tzagoloff for kindly allowing the use of the determined primary sequence of yeast mitochondrial leucyl-tRNA synthetase, Wen-Chy Chu (Jack Horowitz, Iowa State University) for determination of the amino terminus of valyl-tRNA synthetase, Michael Härtlein (Reuben Leberman, European Molecular Biology Laboratory) for sharing his data prior to publication, Ronald C. Wek for helpful discussions and Kurt Gish, Elaine Ito, and Jeanne Sameshima for excellent technical assistance.

REFERENCES

1. Neidhardt, F. C., Parker, J., and McKeever, W. G. (1975) *Annu. Rev. Microbiol.* **29,** 215–250
2. Schimmel, P. R., and Söll, D. (1979) *Annu. Rev. Biochem.* **48,** 601–648
3. Jasin, M., Regan, L., and Schimmel, P. (1984) *Cell* **36,** 1089–1095
4. Putney, S. D., and Schimmel, P. (1981) *Nature* **291,** 632–635
5. Butler, J. S., Springer, M., Dondon, J., and Grunberg-Manago, M. (1986) *J. Bacteriol.* **165,** 198–203
6. Moras, D., Comarmond, M. B., Fischer, J., Weiss, R., Thierry, J. C., Ebel, J. P., and Giegé, R. (1980) *Nature* **288,** 669–674
7. Jordana, X., Chatton, B., Paz-Weisshaar, M., Buhler, J.-M., Cramer, F., Ebel, J. P., and Fasiolo, F. (1987) *J. Biol. Chem.* **262,** 7189–7194
8. Sanger, F., Nicklen, S., and Coulsen, A. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5463–5467
9. Henikoff, S. (1984) *Gene (Amst.)* **28,** 351–359
10. Anderson, S. (1981) *Nucleic Acids Res.* **9,** 3015–3027
11. Biggin, M. D., Gibson, T. J., and Hong, G. F. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80,** 3963–3965
12. Hellings, R. B., Goodman, H. M., and Boyer, H. W. (1974) *J. Virol.* **14,** 1235–1244
13. Yanisch-Perron, C., Vieira, J., and Messing, J. (1985) *Gene (Amst.)* **33,** 103–119
14. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) in *Molecular Cloning: A Laboratory Manual,* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
15. Gross, B. (1985) *DNA Analysis Program for the Macintosh® Computer,* Textco, West Lebanon, New Hampshire
16. Kristofferson, D. (1987) *Nature* **325,** 555–556
17. Devereux, J., and Haeberli, P. (1983) Program Library of the University of Wisconsin Genetics Computer Group, Madison, Wisconsin
18. Clarke, L., and Carbon, J. (1976) *Cell* **9,** 91–99
19. Heck, J. D., and Hatfield, G. W. (1987) *J. Biol. Chem.* **262,** 857–867
20. Yaniv, M., and Gros, F. (1969) *J. Mol. Biol.* **44,** 1–15
21. Konigsberg, W., and Godson, G. N. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80,** 687–691
22. Kern, D., Giegé, R., Robbe-Saul, S., Boulanger, Y., and Ebel, J. P. (1975) *Biochimie* **57,** 1167–1176
23. Schimmel, P., Jasin, M., and Regan, L. (1984) *Fed. Proc.* **43,** 2987–2990
24. Deleted in proof
25. Hoben, P., Royal, N., Cheung, A., Yamao, F., Biemann, K., and Söll, D. (1982) *J. Biol. Chem.* **257,** 11644–11650
26. Breton, R., Sanfaçon, H., Papayannopoulos, I., Biemann, K., and Lapointe, J. (1986) *J. Biol. Chem.* **261,** 10610–10717
27. Webster, T. A., Gibson, B. W., Keng, T., Biemann, K., and Schimmel, P. (1983) *J. Biol. Chem.* **258,** 10637–10641
28. Freedman, R., Gibson, B., Donovan, D., Biemann, K., Eisenbeis, S., Parker, J., and Schimmel, P. (1985) *J. Biol. Chem.* **260,** 10063–10068
29. Webster, T., Tsai, H., Kula, M., Mackie, G., and Schimmel, P. (1984) *Science* **226,** 1315–1317
30. Barker, D. G., Ebel, J. P., Jakes, R., and Burton, C. J. (1982) *Eur. J. Biochem.* **127,** 449–457
31. Fayat, G., Mayaux, J.-F., Sacerdot, C., Fromant, M., Springer, M., Grunberg-Manago, M., and Blanquet, S. (1983) *J. Mol. Biol.* **171,** 239–261
32. Härtlein, M., Madern, D., and Leberman, R. (1987) *Nucleic Acids Res.* **15,** 1005–1017
33. Mayaux, J.-F., Fayat, G., Fromant, M., Springer, M., Grunberg-Manago, M., and Blanquet, S. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80,** 6152–6156
34. Hall, C. V., Van Cleemput, M., Muench, K. H., and Yanofsky, C. (1982) *J. Biol. Chem.* **257,** 6132–6136
35. Barker, D. G., Burton, C. J., and Winter, G. (1982) *FEBS Lett.* **150,** 419–423
36. Tzagoloff, A., Akai, A., Kurkulos, M., and Repetto, B. (1987) *J. Biol. Chem.* **262,** 850–856
37. Jasin, M., Regan, L., and Schimmel, P. (1983) *Nature* **306,** 441–447
38. Zelwer, C., Risler, J. L., and Brunie, S. (1982) *J. Mol. Biol.* **155,** 63–81
39. Wetzel, R. (1978) *Origins Life* **9,** 39–50
40. Blow, D. M., Bhat, T. N., Metcalfe, A., Risler, J. L., Brunie, S., and Zelwer, C. (1983) *J. Mol. Biol.* **171,** 571–576
41. Hountondji, C., Dessen, P., and Blanquet, S. (1986) *Biochimie* **68,** 1071–1078
42. Doolittle, R. F. (1981) *Science* **214,** 149–159
43. Berg, J. M. (1986) *Science* **232,** 485–487
44. Kohda, D., Yokoyama, S., and Miyazawa, T. (1984) *FEBS Lett.* **174,** 20–23
45. Winter, G., Fersht, A. R., Wilkinson, A. J., Zoller, M., and Smith, M. (1982) *Nature* **299,** 756–758
46. Doolittle, R. F. (1986) in *Of Urfs and Orfs,* University Science Books, Mill Valley, CA

Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 2441–2445, March 1995
Evolution

Root of the universal tree of life based on ancient aminoacyl-tRNA synthetase gene duplications

(Archaea/Bacteria/eukaryote/phylogeny)

JAMES R. BROWN* AND W. FORD DOOLITTLE

Canadian Institute for Advanced Research, Department of Biochemistry, Dalhousie University, Sir Charles Tupper Building, Halifax, NS, Canada B3H 4H7

Communicated by Carl R. Woese, University of Illinois, Urbana, IL, November 8, 1994 (received for review October 1, 1994)

ABSTRACT **Universal trees based on sequences of single gene homologs cannot be rooted. Iwabe *et al.* [Iwabe, N., Kuma, K.-I., Hasegawa, M., Osawa, S. & Miyata, T. (1989) *Proc. Natl. Acad. Sci. USA* 86, 9355–9359] circumvented this problem by using ancient gene duplications that predated the last common ancestor of all living things. Their separate, reciprocally rooted gene trees for elongation factors and ATPase subunits showed Bacteria (eubacteria) as branching first from the universal tree with Archaea (archaebacteria) and Eucarya (eukaryotes) as sister groups. Given its topical importance to evolutionary biology and concerns about the appropriateness of the ATPase data set, an evaluation of the universal tree root using other ancient gene duplications is essential. In this study, we derive a rooting for the universal tree using aminoacyl-tRNA synthetase genes, an extensive multigene family whose divergence likely preceded that of prokaryotes and eukaryotes. An approximately 1600-bp conserved region was sequenced from the isoleucyl-tRNA synthetases of several species representing deep evolutionary branches of eukaryotes (*Nosema locustae*), Bacteria (*Aquifex pyrophilus* and *Thermotoga maritima*) and Archaea (*Pyrococcus furiosus* and *Sulfolobus acidocaldarius*). In addition, a new valyl-tRNA synthetase was characterized from the protist *Trichomonas vaginalis*. Different phylogenetic methods were used to generate trees of isoleucyl-tRNA synthetases rooted by valyl- and leucyl-tRNA synthetases. All isoleucyl-tRNA synthetase trees showed Archaea and Eucarya as sister groups, providing strong confirmation for the universal tree rooting reported by Iwabe *et al.* As well, there was strong support for the monophyly (*sensu* Hennig) of Archaea. The valyl-tRNA synthetase gene from *Tr. vaginalis* clustered with other eukaryotic ValRS genes, which may have been transferred from the mitochondrial genome to the nuclear genome, suggesting that this amitochondrial trichomonad once harbored an endosymbiotic bacterium.**

Studies of early cellular evolution have been greatly influenced by two major findings of molecular systematics. First was the revelation from phylogenetic analyses of rRNA molecules that the universal tree of life consists of three domains: the Archaea (archaebacteria), Bacteria (eubacteria), and Eucarya (eukaryotes) (1, 2). Second was the reciprocal rooting of gene trees for two separate paralogous gene families—the genes encoding elongation factors (EFs) Tu/1α and G/2 (3) and the ATPase α and β subunits (3, 4)—which showed that Archaea and Eucarya are sister groups.

Despite the recent expansive growth of gene data bases, no other paralogous gene phylogenies have been developed that might allow us to confirm the root of the universal tree. The phylogenies of several other macromolecules, including RNA polymerases (5) and many ribosomal proteins (6), are indeed consistent with the subdivision of life into three domains, with archaeal and eukaryotic gene homologs being least distant from each other. However, such single gene trees cannot be rooted, and thus the closeness of archaeal and eukaryotes may simply mean that their genes mutate more slowly than do those of bacteria.

New gene discoveries and recent critiques have cast some significant doubt on the validity of conclusions based on duplicated genes for EF and ATPase subunits. Recently, V-type-like ATPases (previously known to exist only in eukaryotes and archaea), similar to archaeal V-type ATPases, have been found in two species of bacteria (7, 8), and a bacterial-like F1- ATPase β-subunit gene has been detected in the archaeon *Methanosacrina barkeri* (9). Collectively, these data suggest that either the full family structure of ATPase-duplicated genes has yet to be determined (10) or that extensive lateral gene transfers between domains have occurred (11), thus rendering any conclusions about domain relationships based on ATPase gene phylogenies suspect. Forterre *et al.* (10) have expressed concerns over the small number of amino acid positions that can be confidently aligned between the EF-Tu/1α and EF-G/2 genes and the paucity of taxa used by Iwabe *et al.* (3). Recent analyses involving a broader species data base, in particular new archaeal EF genes, produce statistically reliable trees using EF-G/2 but not EF-Tu/1α sequences (12). Therefore, the rooting of the universal tree remains an important question that must be addressed not only through a reanalysis of existing EF and ATPase data but also by using other ancient duplicated gene families.

One such promising duplicated gene family comprises the aminoacyl-tRNA synthetases, which catalyze the esterification or "charging" of a single amino acid to its cognate tRNA molecule. The function and structure of aminoacyl-tRNA synthetases have been intensely studied, especially with respect to mechanisms of amino acid charging and tRNA specificity (ref. 13; reviewed in ref. 14). On the basis of sequence similarity and crystallographic structure, aminoacyl-tRNA synthetases are classified as being either group I (specific for glutamic acid, glutamine, tryptophan, tyrosine, valine, leucine, isoleucine, methionine, cysteine, and arginine) or group II (specific for threonine, proline, serine, lysine, aspartic acid, asparagine, histidine, alanine, glycine, and phenylalanine). Group I aminoacyl-tRNA synthetase all share two consensus amino acid motifs, "HIGH" (His-Ile-Gly-His) and "KMSKS" (Lys-Met-Ser-Lys-Ser), while group II synthetases lack these motifs but have a third consensus region "GLER" (Gly-Leu-Glu-Arg). Despite having similar catalytic function, groups I and II aminoacyl-tRNA synthetases do not appear to be related in sequence or higher order structure.

Nagel and R. Doolittle (15) showed that all aminoacyl-tRNA synthetases within a specific group (I or II) are related and that bacterial and eukaryotic versions of aminoacyl-tRNA

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: IleRS, LeuRS, and ValRS, isoleucyl-, leucyl-, and valyl-tRNA synthetases, respectively; EF, elongation factor.

*To whom reprint requests should be addressed.

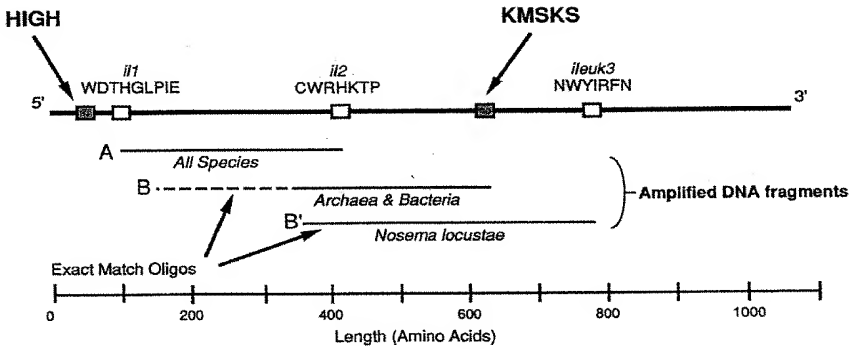


FIG. 1. Schematic diagram of PCR amplification and cloning strategy for IleRS genes. The length of IleRS gene products is known to vary from 939 to 1081 amino acids (40). "HIGH" and "KMSKS" are amino acid motifs conserved in all group I aminoacyl-tRNA synthetase gene products. Initially, an approximately 950-bp region of IleRS, corresponding to fragment A, was amplified from all species using the degenerate oligonucleotide primers *il1* and *il2*. Fragment A was subsequently cloned and sequenced in its entirety. Later, a second round of amplifications was done with a species-specific primer designed to anneal within fragment A and a degenerate primer designed to the KMSKSLGN motif, which generated fragment B. The exception was *N. locustae* where the primer *ileuk3* replaced the *KMSKS* primer to amplify the fragment B'. The fragments A and B or B' overlapped, producing approximately 1600 bp of continuous sequence for all species (1884 bp for *N. locustae*).

synthetases charging a particular amino acid always cluster together to the exclusion of synthetases recognizing other amino acids. Their separate phylogenetic trees for group I and II aminoacyl-tRNA synthetases suggest that the amino acid-specific synthetases are ancient proteins that diverged prior to the emergence of prokaryotic and eukaryotic lineages. Thus, it is reasonable to attempt to root a universal tree derived from one amino acid type of group I aminoacyl-tRNA synthetase with the sequences of another group I aminoacyl-tRNA synthetase.

In the present study, nearly the entire region between the HIGH and KMSKS motifs (about 1600 bp in length) was cloned and sequenced from the group I isoleucyl-tRNA synthetase (IleRS) gene of several lower eukaryotes, bacteria, and archaea. This portion of the gene represents the most conserved region, both within and between different types of group I aminoacyl-tRNA synthetases. The aminoacyl-tRNA synthetases for three aliphatic amino acids (valine, leucine, and isoleucine) were chosen because (*i*) these synthetases appear (15) to be most the recently diverged (which facilitates their alignment) and (*ii*) prior to this study, IleRS was the only aminoacyl-tRNA synthetase characterized from an archaeon, the methanogen *Methanobacterium thermoautotrophicum* (16). In this report, additional archaeal IleRS sequences were determined from the species *Pyrococcus furiosus* (like *M. thermoautotrophicum*, a member of the Euryarchaeota) and *Sulfolobus acidocaldarius*, a member of the Crenarchaeota. IleRS sequences were also determined for species which, according to rRNA phylogenies, are among the most deeply branching lineages of Bacteria [*Aquifex pyrophilus* (17) and *Thermotoga maritima* (18)] and eukaryotes (*Nosema locustae*, an amitochondrial microspordian). As well a new ValRS from the early-diverging eukaryote *Trichomonas vaginalis*, was sequenced.†

By rooting the IleRS gene tree with ValRS and LeuRS genes, our analysis provides significant, independent collaboration of the earlier conclusions of Iwabe *et al.* about the close relationship between Archaea and eukaryotes and the bacterial root of the universal tree. Furthermore, the three domains are shown to be separate monophyletic groups, a finding that is incompatible with the eocyte hypothesis of eukaryotic origins (19).

†Sequences reported in this paper have been deposited in the GenBank data base (accession nos. L37096–L37098 and L37104–L37106).

MATERIALS AND METHODS

DNA Sources. Genomic DNA samples were gifts from the following individuals: *A. pyrophilus* from R. Huber (Universität Regensburg, Germany), *P. furiosus* from F. Robb (University of Maryland, Baltimore), *Th. maritima* from P. Dennis (University of British Columbia, Vancouver, Canada), *Tr. vaginalis* from M. Müller (Rockefeller Institute, New York), and *N. locustae* from A. Roger (this laboratory) prepared from spores obtained from ATCC (no. 30860). *S. acidocaldarius* genomic DNA was prepared from laboratory cultured cells (a gift from W. Zillig, Max-Planck-Institut für Biochemie, Martinsried, Germany). Other DNA sequences were obtained from public data bases.

PCR Amplification, Cloning, and Sequencing. An approximately 1600-bp region of the IleRS genes from *A. pyrophilus, P. furiosus, S. acidocaldarius*, and *Th. maritima* and a 1900-bp region from the *N. locustae* gene were PCR-amplified with two sets of oligonucleotide primers (Fig. 1). The first set of primers was designed with partial degeneracy to the amino sequences Trp-Asp-Thr-His-Gly-Leu-Pro-Ile-Glu (WDTAGLPIE in single-letter code in Fig. 1) (5′-TGGGAYACNCAYGGNYT-NCCNRTNGA-3′ named *il1*) and Cys-Trp-Arg-(His or Cys or Ser)-(Lys or Asp)-Thr-Pro (CWRHKTP in single-letter code in Fig. 1) (complement 5′-GGNGTNTYRCWNCKCCAR-CA-3′ named *il2*). This primer pair consistently amplified a fragment about 950 bp long in the tested species. In a second PCR experiment, the remaining portion of the gene was amplified by using a species-specific 5′-end primer (primer sequences available upon request from J.R.B.) designed to anneal within the *il1*/*il2*-cloned fragment and a complementary degenerate primer designated *KMSKS* designed to the amino acid motif Lys-Met-Ser-Lys-Ser-Leu-Gly-Asn (KMSK-SLGN in single-letter code) (5′-RTTWCCHARWSAYT-TWSACATYTT-3′). For *N. locustae*, the *KMSKS* primer was replaced by the primer *ileuk3*, complementary to the amino acid sequence Asn-Trp-Tyr-Ile-Arg-Phe-Asn (NWYIRFN in single-letter code in Fig. 1) (5′-RTTNARNCKDATRTAC-CARTT-3′) located about 300 bp downstream of the KMSK-SLGN motif. A 1480-bp region of the ValRS gene comparable to the IleRS 1600-bp section was amplified from *Tr. vaginalis* by using a ValRS-specific primer *val1*, which matches the 5′ end amino acid sequence Asp-His-Ala-Gly-Ile-Ala-Thr-Glu (DHAGIATQ in single-letter code) (5′-GAYCAYGCWGG-WATWGCWACNCA-3′) and the *KMSKS* primer.

Thermal cycle amplifications were performed in 50-μl final volume with 5 μl of 10× reaction buffer (500 mM KCl/100 mM Tris·HCl, pH 8.3/15 mM $MgCl_2$/0.1% gelatin) containing

dNTPs at 200 mM, primers at 5–7 μM, sample DNA (≈50 ng), and 0.5 units of *Thermus aquaticus* DNA polymerase with 50 μl of mineral oil overlaid. The reaction cycles consisted of denaturation for 1 min at 95°C, primer annealing for 1 min at 48°C, and extension for 2 min at 72°C. Cycles were repeated 40 times, and the final cycle included an extension reaction of 5 min. Negative controls (all of the above reagents except for template DNA) were included in all amplification series as a screen for possible foreign DNA contamination.

Amplified DNA samples were electrophoresed in 2.0% low-melting-point agarose gels in separate gel apparatuses, and the fragments were extracted either by the phenol method (20) or with the Prep-a-Gene kit according to the vender's protocols (Bio-Rad). Isolated DNA fragments were then subcloned into the pCRII vector by following the vendor's methods (Invitrogen). Double-stranded DNA was sequenced by using the dideoxynucleotide chain-termination method (21) and T7 polymerase (United States Biochemical) and following standard protocols. One DNA strand was sequenced in its entirety, and, depending on the species, about 40–80% of the complementary strand was also determined by using internal oligonucleotide primers. All ambiguous regions were confirmed by sequencing the opposite strand.

Sequence Alignments. New sequences were edited by using the program ESEE (22). IleRS, LeuRS, and ValRS sequences were obtained from National Center for Biotechnology Information data base by using Network ENTREZ software. Amino acid sequences were first aligned with the program MULTALIN (23) and then edited by eye to better align certain conserved motifs. The final alignment was in good agreement with those done previously (15, 24). Since the placement of some gaps is variable, all insertions/deletions were edited from multiple sequence alignments, leaving 354 amino acid positions for the phylogenetic analysis.

Phylogenetic Analyses. Phylogenetic trees were constructed by using both maximum parsimony and distance methods. Maximum parsimony analysis was done with the software packages PAUP version 3.1.1 (25) and PHYLIP version 3.5 (26). The large size of this data set did not permit an exhaustive search for the total number of minimal-length trees. Instead, the program PAUP was used to estimate the number and length of minimal trees from 20 replicate random heuristic searches with the PROTPARS stepmatrix to specify the minimum number of nucleotide replacements required to change from one amino acid to another. The programs SEQBOOT, PROTPARS, and CONSENSE of the PHYLIP 3.5 package were used to derive confidence limits, estimated by 300 bootstrap-replicates, for branch points in the maximum parsimony tree.

A distance matrix of pairwise comparisons of the proportion of different amino acids per site was constructed by using the program PROTDIST (26). In our analysis, we invoked the "Dayhoff" program option, which estimates the expected amino acid replacements per position by using a replacement model based on the Dayhoff 120 matrix. The programs SEQBOOT, NEIGHBOR, and CONSENSE were used to derive a neighbor-joining tree with confidence limits estimated by 300 bootstrap replications.

RESULTS

Sequence Analysis. The five new IleRS and one ValRS sequence shared many similarities with known aminoacyl-tRNA synthetases of their respective type. Conserved sequence motifs previously noted in IleRS genes were found in all five new IleRS sequences. However, some species had unique insertions in different regions of the molecule, ranging in length from as few as 1 to as many as 33 amino acids. None of these insertions were concordant synapomorphies among species. All gaps in the alignment were omitted, which left 354 amino acid positions for the phylogenetic analysis.

For IleRS sequences, mean intradomain sequence identity values were 57% for archaea, 54% for bacteria, and 61% for eukaryotes. As expected, mean sequence identity comparisons between archaea and eukaryotes (41%), archaea and bacteria (45%), or eukaryotes and bacteria (35%) were lower (sequence alignment and pairwise distance comparisons are available upon request from J.R.B.).

Phylogenetic Analyses. Only one minimal length tree was found after 20 random replicates of maximum parsimony analysis using the PROTPARS stepmatrix in the program PAUP. This tree was 3512 steps long and showed archaea and eukaryotes as sister groups. This grouping occurred in 75% of 300 bootstrap replicates of maximum parsimony analysis using the program PROTPARS (Fig. 2). A similar tree topology was obtained with a simple progressive scalar scheme for down-weighting increasingly variable sites in maximum parsimony analysis (implemented in PAUP). Bootstrap analysis of this weighted parsimony method showed 70% support for the Archaea–Eucarya clade. Unweighted parsimony also recovered the Archaea–Eucarya clade in the minimal length tree, although bootstrap analysis resulted in low statistical support (about 60%) for an Archaea–Bacteria clade.

The distance method, using expected amino acid replacements per site (calculated by the program PROTDIST) to construct a neighbor-joining tree, also supported the Archaea–Eucarya clade at a high bootstrap value of 85% (Fig. 3). All phylogenetic analyses consistently supported, with bootstrap confidence limits ranging from 88% to 100%, the separate monophyletic groups of archaea, bacteria, and eukaryotes.

Within domains, the branching order of individual species was less well resolved. Both phylogenetic methods separated the archaeal groups, Euryarchaeota and Crenarchaeota, although with low bootstrap confidence limits. Although the internal nodes are not statistically significant, the maximum parsimony tree appears to agree best with the expectations of branching order proposed by rRNA phylogenies for within eukaryotes (*N. locustae* being the lowest branch) and bacteria

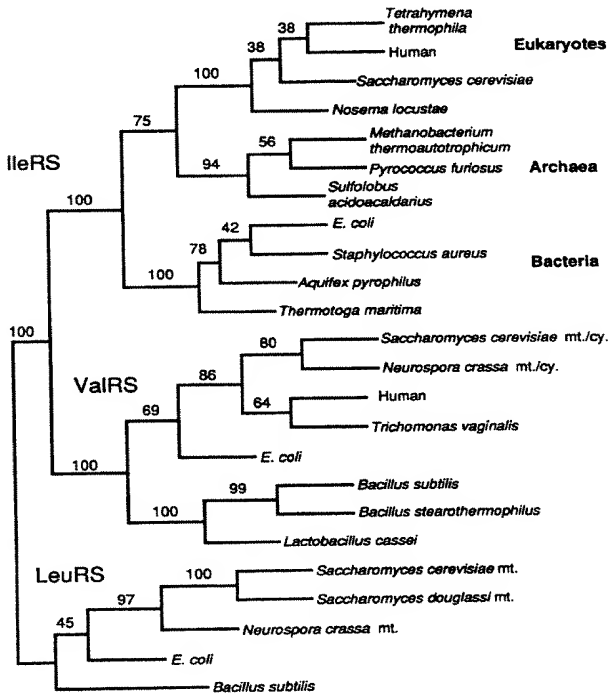


FIG. 2. Consensus maximum parsimony tree of IleRS, ValRS, and LeuRS genes using the program PROTPARS (26). Numbers are the frequency of occurrence of nodes in 300 bootstrap replicates.

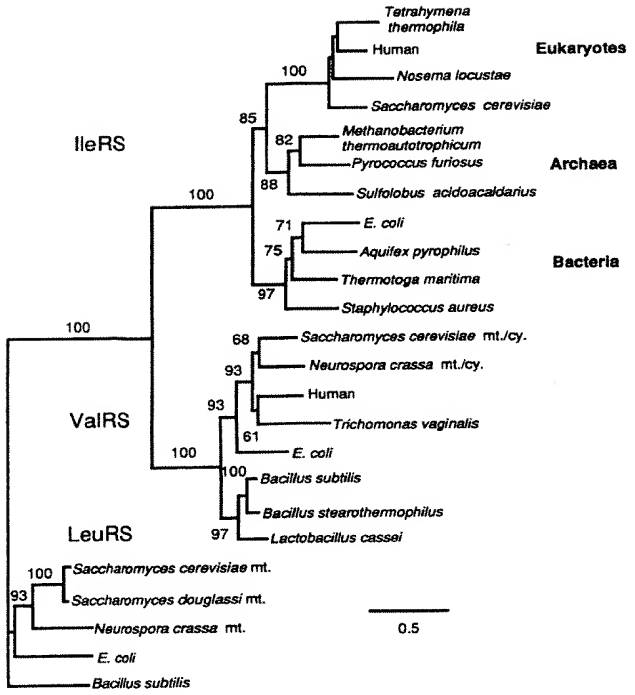


FIG. 3. Neighbor-joining tree of IleRS, ValRS, and LeuRS genes using the program NEIGHBOR (26). The scale represents 0.5 expected number of amino acid replacements per position as determined with the program PROTDIST. Numbers are the frequency of occurrence of nodes that exceeded 50% of 300 bootstrap replicates.

(*Th. maritima* and *A. pyrophilus* nearest the root). Better resolution of taxa within the IleRS phylogeny can likely be obtained by using the full-length sequences rather than only those positions that can be confidently aligned with ValRS and LeuRS (J.R.B., unpublished data).

The product of ValRS genes of the yeast *Neurospora* and humans are utilized in both the cytoplasm and mitochondria, so the placement of *E. coli* ValRS at the root of the eukaryotes suggests that nuclear copies of these ValRS genes may have originated from a mitochondrial endosymbiont. Thus, the firm placement of the amitochondrial protist *Tr. vaginalis* with the rest of the eukaryote mitochondrial isoforms with high bootstrap confidence limits is surprising.

DISCUSSION

Maximum parsimony and neighbor-joining distance trees both show that (*i*) the three sets of aminoacyl-tRNA synthetase genes form monophyletic groups, in agreement with the analysis of Nagel and Doolittle (15) of the entire group I family of genes; (*ii*) within the IleRS portion of the tree, Archaea, Bacteria, and Eucarya are separate monophyletic domains; and (*iii*) Archaea and eukaryotes are supported as a clade according to heuristic search methods for the minimal-length tree as well as bootstrap analysis, which, under most conditions, is considered to be a conservative estimate of the significance of branching points (27).

The IleRS tree provides important confirmation of the rooting of the universal tree in the lineage leading to the bacteria as suggested by the analysis of Iwabe *et al.* (3) of the duplicated genes encoding EF-Tu/1α and EF-G/2. The EF gene analysis involved the reciprocal rooting of two gene trees, both of which included representative species from all three domains. In our study, only the IleRS gene tree has a full complement of archaeal, bacterial, and eukaryotic species, since ValRS and LeuRS genes are unknown for the archaea. However, we consider our result to be the strongest to date in support of the sisterhood of archaea and eukaryotes. The present IleRS data set exceeds that of the EF-Tu/1α gene family in terms of sequence length—354 amino acids for IleRS versus 120 amino acids for the joint EF alignment. Furthermore, the IleRS dataset includes more deeply branching species within the eukaryotes and bacteria and a more comprehensive selection of archaea.

The analysis of duplicated genes performed by Iwabe *et al.* (3) involved only single archaeal homologs and thus did not address the issue of the coherence of the Archaea. Monophyletic groupings of archaeal, bacterial, and eukaryotic clades are strongly supported by the present phylogenetic analysis. In confirming a root between bacteria and archaea/eukaryotes, the IleRS data set also supports inferences concerning the monophyly of each domain based on unrootable data (for instance the rRNA sequences) and are inconsistent with treatments of this data that would place the root between the Euryarchaeotes and Crenarchaeotes [as in the Lake 1988 version of the "eocyte tree" (19)].

The congruence of IleRS and EF gene trees is not surprising, given that aminoacyl-tRNA synthetases and EF Tu/1α sequentially interact with the tRNA-amino acid complex and, as such, might have coevolved functions. The greater similarity of the archaea to eukaryotes rather than to bacteria is supported by several lines of evidence involving the cell's genetic machinery. These include recent findings of archaeal homologs to eukaryotic TATA-binding proteins (28, 29), transcription factor TFIIB (30), and a TFIIS-like sequence (31) as well as the closer sequence similarity of genes for RNA polymerase (5) and many ribosomal proteins (6).

Other data sets sometimes suggest alternative relationships between the three domains. For example, glutamine synthetase trees place archaea and the Gram-positive bacteria in the same clade: some sort of lateral transfer might be the best explanation (32). Similarly, some eukaryotic nuclear genes—in addition to those likely derived from mitochondrial or plastid genomes—appear of bacterial rather than archaeal origin. (An example is phosphoglycerate kinase). Other authors (33, 34) have claimed that such occurrences bespeak a radical chimerism, the eukaryotic nucleus for instance being the product of the fusion of the entire genomes of archaea and bacteria. Although the present data do not address these issues directly, they add to a considerable body of evidence in favor of the notion that the eukaryotic transcription and translation machinery—surely the core of cell biology—are archaeal in nature. Whether other nuclear genes of apparent bacterial origin were acquired in some genetic cataclysm or one-by-one over hundreds of millions of years remains an open question.

While eukaryotes must have a full suite of aminoacyl-tRNA synthetases that are functional in both the cytoplasm and the mitochondria, the mode of coding for specific cellular isoforms varies with amino acid type. For example, there are two separate LeuRS genes coding for cytoplasmic and mitochondrial isoforms (35), while the same ValRS gene product is used in both cellular locations (36–38). Our analysis suggests that eukaryotic cytoplasmic IleRS are the product of ancient nuclear genes, while the single eukaryotic ValRS may have been of bacterial (endosymbiotic mitochondrial) origin. Thus, the placement of the amitochondrial protist, *Tr. vaginalis*, with the remaining eukaryotes is a surprising result and suggests that the nuclear genome of *Tr. vaginalis* may have experienced a similar introduction of certain genes from an endosymbiont. While trichomonads lack a mitochondrion, they do have another organelle, the hydrogenosome, which may have had an endosymbiotic origin (reviewed in ref. 39). However, any conclusions about the relationships among ValRS genes must remain highly speculative, given the limited number of genes known from lower eukaryotes. Furthermore, archaeal se-

quences would be essential for determining the exact topology of the ValRS portion of the tree.

Our analysis opens several new avenues of research. Group I, as well as group II aminoacyl-tRNA synthetases are large multigene families that conceivably offer several other opportunities for testing the root of the universal tree. Although group I and II aminoacyl-tRNA synthetases have the same catalytic function, the two gene families are highly divergent at the amino acid sequence level and appear to employ different modes of tRNA recognition (14, 40). Given the evolutionary distinctiveness of the two groups of aminoacyl-tRNA synthetases, it has been postulated that, at one time, there may have been two independent protein synthetic systems working with reduced sets of amino acids that subsequently merged into the present-day genetic code (15). The degree of congruence between group I and II aminoacyl-tRNA synthetase gene trees using new archaeal sequences might provide further insight into the level of refinement of the genetic machinery of the last common ancestral cell.

We thank Mark Ragan of the National Research Council for assistance in running the PHYLIP software, Jeremy Murray for some technical assistance, and all the individuals who generously provided genomic DNA samples. W.F.D. is a Fellow of the Canadian Institute for Advanced Research, and J.R.B. is a Canadian Medical Research Council (M.R.C.) Postdoctoral Fellow. This work was supported by an operating grant from M.R.C. to W.F.D.

1. Woese, C. R. & Fox, G. E. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5088–5090.
2. Woese, C. R., Kandler, O. & Wheelis, M. L. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 4576–4579.
3. Iwabe, N., Kuma, K.-I., Hasegawa, M., Osawa, S. & Miyata, T. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 9355–9359.
4. Gogarten, J. P., Kibak, H., Dittrich, P., Taiz, L., Bowman, E. J., Bowman, B. J., Manolson, N. F., Poole, R. J., Date, T., Oshima, T., Konishi, J., Denda, K. & Yoshida, M. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 6661–6665.
5. Pühler, G., Leffers, H., Gropp, F., Palm, P., Klenk, H.-P., Lottspeich, F., Garrett, R. A. & Zillig, W. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 4569–4573.
6. Ramirez, C., Köpke, A. K. E., Yang, D.-C., Boeckh, T. & Matheson, A. T. (1993) *The Biochemistry of Archaea (Archaebacteria)* eds. Kates, M., Kushner, D. J. & Matheson, A. T. (Elsevier, Amsterdam), pp. 439–466.
7. Tsutsumi, S., Denda, K., Yokoyama, K., Oshima, T., Date, T. & Yoshida, M. (1991) *Biochim. Biophys. Acta* **1098,** 13–20.
8. Kakinuma, Y., Igarishi, K., Konishi, K. & Yamato, I. (1991) *FEBS Lett.* **292,** 64–68.
9. Sumi, M., Sato, M. H., Denda, K., Date, T. & Yoshida, M. (1992) *FEBS Lett.* **314,** 207–210.
10. Forterre, P., Benachenhou-Lahfa, N., Confalonieri, F., Duguet, M., Elie, C. & Labedan, B. (1993) *BioSystems* **28,** 15–32.
11. Hilario, E. & Gogarten, J. P. (1993) *BioSystems* **31,** 111–119.
12. Creti, R., Ceccaralli, E., Bocchetta, M., Sanangelantoni, A. M., Tiboni, O., Palm, P. & Cammarano, P. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 3255–3259.
13. Eriani, G., Delarue, M., Poch, O., Gangloff, J. & Moras, D. (1990) *Nature (London)* **347,** 203–206.
14. Carter, C. W. (1993) *Annu. Rev. Biochem.* **62,** 715–748.
15. Nagel, G. M. & Doolittle, R. F. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 8121–8125.
16. Jenal, U., Rechsteiner, T., Tan, P.-Y., Bühlmann, E., Meile, L. & Leisinger, T. (1991) *J. Biol. Chem.* **266,** 10570–10577.
17. Burggraf, S., Olsen, G. J., Stetter, K. O. & Woese, C. R. (1992) *Sys. Appl. Microbiol.* **15,** 352–356.
18. Achenbach-Richter, L., Gupta, R., Stetter, K. O. & Woese, C. R. (1987) *Sys. Appl. Microbiol.* **9,** 34–39.
19. Lake, J. A. (1988) *Nature (London)* **331,** 184–186.
20. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 6.30–6.31.
21. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
22. Cabot, E. L. & Beckenbach, A. T. (1989) *Comput. Appl. Biosci.* **5,** 233–234.
23. Corpet, F. (1988) *Nucleic Acids Res.* **16,** 10881–10890.
24. Shiba, K. & Schimmel, P. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 1880–1884.
25. Swofford, D. L. (1993) PAUP: *Phylogenetic Analysis Using Parsimony* (Illinois Nat. Hist. Survey, Champaign, IL), Version 3.1.1.
26. Felsenstein, J. (1993) PHYLIP *Phylogeny Inference Package* (Dept. of Genet., Univ. of Washington, Seattle), Version 3.5c.
27. Hillis, D. M. & Bull, J. J. (1993) *Syst. Biol.* **42,** 182–192.
28. Rowlands, T., Baumann, P. & Jackson, S. P. (1994) *Science* **264,** 1326–1329.
29. Marsh, C. L., Reich, C. I., Whitlock, R. B. & Olsen, G. J. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 4180–4184.
30. Ouzounis, C. & Sander, C. (1992) *Cell* **71,** 189–190.
31. Kaine, B. P., Mehr, I. J. & Woese, C. R. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 3854–3856.
32. Brown, J. R., Masuchi, Y., Robb, F. T. & Doolittle, W. F. (1994) *J. Mol. Evol.* **38,** 566–576.
33. Sogin, M. L. (1991) *Curr. Opin. Genet. Dev.* **1,** 457–463.
34. Zillig, W. (1991) *Curr. Opin. Genet. Dev.* **1,** 544–551.
35. Chow, C. M. & RajBhandary, U. L. (1989) *Mol. Cell. Biol.* **9,** 4645–4652.
36. Jordana, X., Chatton, B., Paz-Weisshaar, M., Buhler, J.-M., Cramer, F., Ebel, J. P. & Fasiolo, F. (1987) *J. Biol. Chem.* **262,** 7189–7194.
37. Hsieh, S.-L. & Campbell, R. D. (1991) *Biochem. J.* **278,** 809–816.
38. Martindale, D. W., Gu, Z. M. & Csank, C. (1989) *Curr. Genet.* **15,** 99–106.
39. Müller, M. (1993) *J. Gen. Microbiol.* **139,** 2879–2889.
40. Cavarelli, J. & Moras, D. (1993) *FASEB J.* **7,** 71–86.

JOURNAL OF BACTERIOLOGY, Apr. 1997, p. 2472–2478
0021-9193/97/$04.00+0

Vol. 179, No. 8

Structure and Regulation of Expression of the *Bacillus subtilis* Valyl-tRNA Synthetase Gene

DONG LUO, JOSETTE LEAUTEY, MARIANNE GRUNBERG-MANAGO, AND HARALD PUTZER*
UPR 9073, CNRS, Institut de Biologie Physico-Chimique, 75005 Paris, France

Received 26 November 1996/Accepted 5 February 1997

We have sequenced the valyl-tRNA synthetase gene (*valS*) of *Bacillus subtilis* and found an open reading frame coding for a protein of 880 amino acids with a molar mass of 101,749. The predicted amino acid sequence shares strong similarity with the valyl-tRNA synthetases from *Bacillus stearothermophilus*, *Lactobacillus casei*, and *Escherichia coli*. Extracts of *B. subtilis* strains overexpressing the *valS* gene on a plasmid have increased valyl-tRNA aminoacylation activity. Northern analysis shows that *valS* is cotranscribed with the *folC* gene (encoding folyl-polyglutamate synthetase) lying downstream. The 300-bp 5′ noncoding region of the gene contains the characteristic regulatory elements, T box, "specifier codon" (GUC), and *rho*-independant transcription terminator of a gene family in gram-positive bacteria that encodes many aminoacyl-tRNA synthetases and some amino acid biosynthetic enzymes and that is regulated by tRNA-mediated antitermination. We have shown that *valS* expression is induced by valine limitation and that the specificity of induction can be switched to threonine by changing the GUC (Val) specifier triplet to ACC (Thr). Overexpression of *valS* from a recombinant plasmid leads to autorepression of a *valS-lacZ* transcriptional fusion. Like induction by valine starvation, autoregulation of *valS* depends on the presence of the GUC specifier codon. Disruption of the *valS* gene was not lethal, suggesting the existence of a second gene, as is the case for both the *thrS* and the *tyrS* genes.

The aminoacyl-tRNA synthetases (aaRS) catalyze the covalent attachment of amino acids to their cognate tRNAs, a reaction crucial for the accuracy of protein synthesis. For the most part, there is only one aaRS for each amino acid species in bacteria, although several exceptions are known. The presence of two very similar lysyl-tRNA synthetases represents the singular exception in *Escherichia coli* (21, 22, 26), where the tRNA synthetases for all 20 amino acids have been cloned (12). The situation is different in gram-positive organisms. On the one hand, they lack a glutaminyl-tRNA synthetase (43), and on the other hand, there are two distinct threonyl-tRNA synthetase genes (*thrS* and *thrZ* [32]) and two tyrosyl-tRNA synthetase genes (*tyrS* and *tyrZ* [9, 20]) in *Bacillus subtilis* and two histidyl-tRNA synthetase genes in *Lactococcus lactis* (36). Chances are that other duplicate genes will be identified with further progress in the various genome-sequencing projects. We have previously shown that the normally silent *thrZ* gene is induced during threonine starvation or by reducing the intracellular concentration of the housekeeping synthetase, ThrS (33).

In contrast to *E. coli*, in which the mechanisms for aaRS gene regulation are as disparate as the number of genes studied (for a review, see references 12 and 34), most of the *B. subtilis* genes isolated appear to be regulated by a common mechanism. Of the 15 tRNA synthetase genes cloned and sequenced in *B. subtilis* (for a review, see references 4 and 34), all but the asparaginyl (*asnS* [2])-, glutamyl (*gltX* [7])-, lysyl (*lysS* [31])-, and methionyl (*metS* [31])-tRNA synthetase genes share common sequence and structural motifs in the leader regions upstream of the translation initiation site (14). Their leader regions are about 300 bp long, and each contains a transcriptional terminator immediately preceded by a 14-nucleotide consensus sequence known as the T box (19, 20, 33). This configuration is found not just in the aaRS genes but also in several of the amino acid biosynthetic operons in *Bacillus* spp. and other gram-positive organisms (13, 34). The leader region of the *thrZ* gene extends over 800 bases and comprises three such tandem domains (33).

For several genes of this family, it has been shown that they are specifically induced by starvation for their cognate amino acid via a mechanism involving transcriptional antitermination. This is the case in the *tyrS* (14), *pheS* (35), and *thrS* and *thrZ* genes (33) and the *ilv-leu* operon (11). *thrS* and *thrZ* are also autorepressed by overproduction of the synthetases themselves (8, 33).

Base pairing between part of the conserved T-box sequence and an equally conserved sequence in the 5′ half of the terminator stem can lead to the formation of an alternative, and mutually exclusive, structure called the antiterminator (14, 33). Studying the *tyrS* system, Grundy and colleagues first provided evidence that the uncharged tRNA can stabilize the formation of the antiterminator structure by interacting with two sites in the leader mRNA (14, 15). The first is between the anticodon loop of the uncharged tRNA and a "specifier codon" that is likely to be bulged out of a large stem-loop structure found in the 5′ half of the leader RNAs of this gene family. The second proposed interaction occurs through base pairing between the NCCA-3′ acceptor end of the uncharged tRNA (including the discriminator base) and the perfectly complementary -UGGN′- sequence in the T box, which is bulged out in the antiterminator conformation (14, 15). Several mutational studies have now been carried out in different systems, and they basically confirm the importance of the specifier codon for the specificity of induction during cognate amino acid starvation (14, 28, 35). Changing the identity of the specifier codon has, in many cases, permitted a switch in the identity of the regulatory amino acid.

The role of the discriminator base in stabilizing the interaction between the acceptor end of the uncharged tRNA and the T box has been studied for *tyrS* (15), *pheS* (35), and *thrS* (35). These reports show that while this interaction is important, the

* Corresponding author. Mailing address: UPR 9073, CNRS, Institut de Biologie Physico-Chimique, 13 rue Pierre et Marie Curie, 75005 Paris, France. Phone: (33)-1-43298226. Fax: (33)-1-40468331. E-mail: putzer@ibpc.fr.

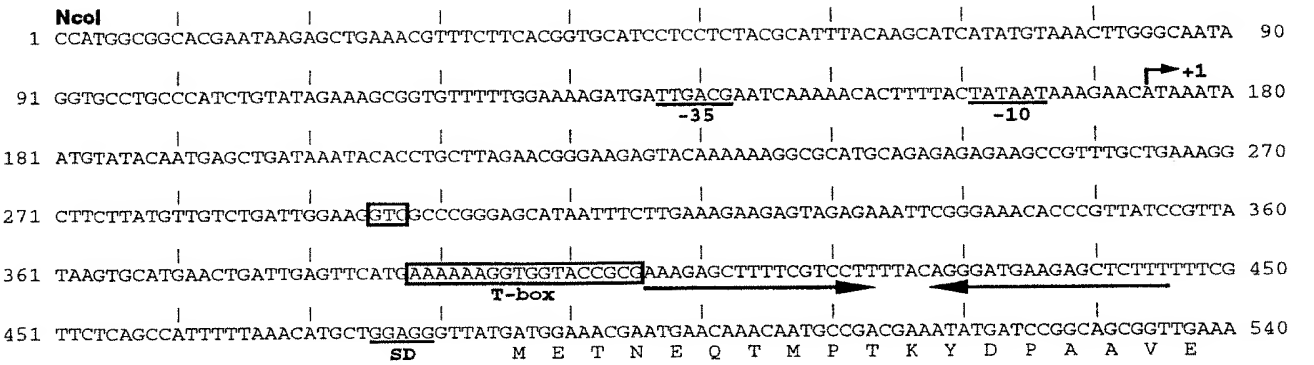


FIG. 1. Nucleotide sequence of the 5′ noncoding region of the *B. subtilis valS* gene. The consensus promoter sequences (−35 and −10 regions) are underlined. The bent arrow indicates the +1 transcription start point. The deduced Shine-Dalgarno-type sequence (SD) is underlined. Converging arrows indicate a potential Rho-independent transcription terminator. The specifier codon (GTC) and the T-box consensus sequence are boxed. The sequence of the whole gene has been deposited in the GenBank/EMBL databases.

sometimes ambiguous results obtained with different mutants suggest that other points of interaction between the tRNA and mRNA, and possibly protein factors, are involved in regulation.

An additional level of regulatory complexity was recently introduced with the discovery that the leader mRNA of *thrS* and at least five other members of this gene family is cleaved just upstream of the transcription terminator in vivo (5). The processed *thrS* transcript is significantly more stable than the full-length mRNA and is the predominant form under threonine starvation conditions. Even though processing can occur in the absence of the tRNA-leader interaction, its contribution to overall induction levels following threonine starvation is substantial (5).

One of the reasons we have studied the expression of the *valS* gene is to find out whether the different aspects of regulation described above apply to other genes of this family or if, on the contrary, some of the regulatory mechanisms are confined to specific genes. For example, all the genes cited above are induced by tRNA-mediated antitermination, but autoregulation has thus far been described only for *thrS* and *thrZ* expression. The only other gene tested in this respect, *pheS*, although induced by phenylalanine starvation, was not repressed by overproduction of phenylalanyl-tRNA synthetase.

In this report, we describe the identification, sequencing and characterization of the *valS* gene. We analyze its transcription pattern and the importance of the specifier codon and the T box for the specificity of induction by valine starvation. Furthermore, we provide evidence that *valS*, like *thrS/thrZ* but unlike *pheS*, is autoregulated.

MATERIALS AND METHODS

Bacterial strains and culture and transformation conditions. All *B. subtilis* strains used in this study are derivatives of the prototrophic strain 168 (BGSC 1A2) or the auxotrophic strain BGSC 1A232 (*ilvD4 trpD2*), containing *valS-lacZ* fusions integrated into the *amy* locus. Strains were grown in M9 minimal medium (29) supplemented with 0.5 mM Trp, 3 mM Ile, 3 mM Leu, 3 mM Val, and trace elements (17). For valine starvation experiments, cells were grown as just described but in the presence of only 0.6 mM Val and harvested for β-galactosidase measurements 2 h after the end of logarithmic growth. Threonine starvation was achieved by the addition of 600 μg of DL-threonine hydroxamate per ml to a M9 medium culture at an optical density at 600 nm of 0.3 to 0.4 (prototrophic strain), which still allowed logarithmic growth. Cells were harvested 2 h later.

Plasmid manipulations were performed in *E. coli* JM109 [*recA1 endA1 gyrA86 thi hsdR17 supE44 relA* λ⁻ Δ(*lac-proAB*), F′(*traD36 proAB lacI*^{q} *lacZ*Δ*M15*)]. *E. coli* KE89 (F⁻ *endA1 hsdR1 hsdM*$^+$ *supE44 thi-1 pcnB*) served as a host for overexpression studies with the *valS*-containing plasmid pHMV11, since this high-copy-number plasmid could not be stably maintained in a *pcnB*$^+$ strain. Concatemeric plasmids for transformation of *B. subtilis* were isolated from *E. coli* JM101 [*thi supE44* Δ(*lac-proAB*) F′(*traD36 proAB*$^+$ *lacI*^{q} *lacZ*Δ*M15*)].

E. coli cells were transformed by electroporation (37), and *B. subtilis* cells were transformed as described elsewhere (25). *E. coli* transformants were selected on LB plates supplemented with 100 μg of ampicillin per ml, and *B. subtilis* transformants were selected on LB plates with 4 μg of chloramphenicol (integrative plasmids) or 10 μg of tetracycline (replicative plasmids) per ml.

Plasmid constructions. Plasmid pDG1129 was a generous gift from P. Stragier. It was constructed by insertion of a 3.15-kb *Nco*I-*Xba*I chromosomal DNA fragment containing the *valS* gene into the vector pMTL22 (3), which was cut with the same enzymes.

For pHMV4, a 1-kb *Bgl*II-*Hin*dIII fragment (coordinates 1 to 978 of the *valS* sequence) from pDG1129 was inserted into plasmid pTZ18R (USB) cut with *Bam*HI and *Hin*dIII.

For pHMV8, the 1-kb insert of pHMV4 was excised as an *Eco*RI-*Hin*dIII fragment and cloned into plasmid pHM2 (8) cut with *Eco*RI and *Hin*dIII.

For pHMV11, the 3.15-kb insert of pDG1129 containing the entire *valS* gene was excised as an *Nsi*I-*Xba*I fragment and inserted into the shuttle vector pHM3 (33) cut with *Pst*I and *Xba*I.

For pHMV12, an internal 1.5-kb *Hin*dIII fragment of *valS* was inserted into the integrative vector pDG641 (16) cut with *Hin*dIII.

For pHMV13, the 1-kb insert of pHMV4 was mutated at two sites: the GUC triplet (coordinates 295 to 297 in Fig. 1) was altered to ACC, and the -TGGT- sequence of the T box (coordinates 396 to 399 in Fig. 1) was changed to -TGGA-. The mutated fragment was excised with *Eco*RI and *Hin*dIII and inserted into pHM2 cut with *Eco*RI and *Hin*dIII.

For pHMV14, the 1-kb insert of pHMV4 where the GUC specifier codon has been mutated to UAA was excised with *Eco*RI and *Hin*dIII and inserted into pHM2 cut with *Eco*RI and *Hin*dIII.

For pHMV15, the 1-kb insert of pHMV4 where the GUC specifier codon has been mutated to ACC was inserted as an *Eco*RI-*Hin*dIII fragment into pHM2 cut with *Eco*RI and *Hin*dIII.

DNA manipulations. The 3.15-kb insert of plasmid pDG1129, containing the entire *valS* gene, was subcloned as three separate fragments in the multicopy plasmids pTZ18R and pTZ19R (USB) for sequencing. The double-stranded recombinant DNAs were used as templates in dideoxy chain termination sequencing reactions (38), using the universal and reverse primers as well as specific synthetic oligonucleotides for the central regions of the cloned fragments.

Site-directed mutagenesis was performed on a single-stranded DNA template by the method of Kunkel et al. (24). Mutations in the *valS* leader were generally introduced on plasmid pHMV4 before being transferred to the *lacZ* fusion vector pHM2. Oligonucleotides used for mutagenesis extended 12 to 15 nucleotides on either side of the mutation site, and sequences are available on request.

RNA manipulations. Total cellular RNA was isolated as described previously (33). Reverse transcriptase assays were carried out with 15 μg of total RNA and about 1 pmol of 5′-end-labeled oligonucleotide (sequence complementary to positions 479 to 498 in Fig. 1). The RNA and oligonucleotide were heated together at 65° for 5 min and then frozen in a mixture of dry ice and ethanol and allowed to thaw on ice. Reactions contained 2 U of avian myeloblastosis virus reverse transcriptase (Eurogentec) and were allowed to run for 30 min at 48°C.

Northern analysis of total cellular RNA was performed as described elsewhere (33). A radiolabeled 1.5-kb *Hin*dIII fragment of the *valS* structural gene (see Fig. 5) was used as a *valS*-specific probe. The *folC* probe was amplified by PCR from chromosomal DNA (positions 73 to 1096 of the structural gene; see Fig. 5).

β-Galactosidase and aminoacylation assays. The β-galactosidase activity of *lacZ* fusions was measured as described previously (33).

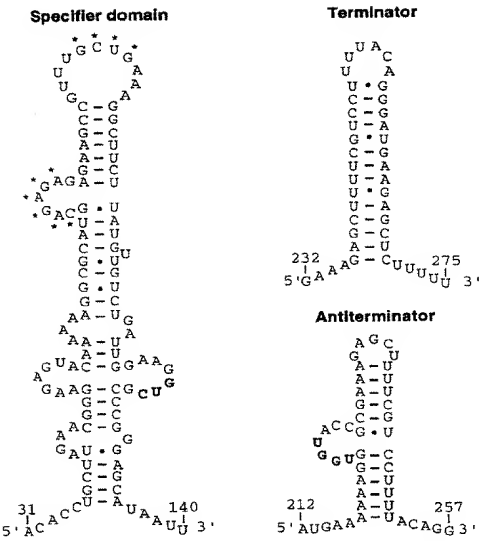


FIG. 2. Putative secondary structures of the specifier domain, the antiterminator, and the terminator of the *B. subtilis valS* leader. The GUC triplet and the -UGGU- sequence in the antiterminator that are believed to interact with the 3′ end of the Val-tRNA^{GAC} are in boldface type. Other conserved sequences (14) are marked by asterisks.

For in vitro aminoacylation measurements, *B. subtilis* or *E. coli* cells harboring recombinant plasmids containing *valS* or the vector alone were grown in LB broth to an optical density at 600 nm of ~1. Cells were harvested and washed with Z buffer without β-mercaptoethanol (29). After suspension in 500 μl of buffer A (10 mM Tris-HCl [pH 7.4], 10% glycerol, 1 mM dithiothreitol), samples were sonicated and clarified by centrifugation. The 100-μl aminoacylation reaction was carried out at 37°C in a reaction mixture containing 50 mM Tris-HCl (pH 7.5), 15 mM $MgCl_2$, 15 mM β-mercaptoethanol, 10 mM ATP, 50 μM L-^{14}C-valine at 200 cpm/pmol, 1 mM dithiothreitol, 120 μg of total *E. coli* tRNA, and various amounts of cellular extract. The nucleic acids were precipitated by trichloroacetic acid and filtered out on GFC filters (Whatman), and the radioactivity retained on the filters was measured by scintillation counting.

Computer analysis. Sequence comparisons were done with the help of the programs BestFit and PileUp of the University of Wisconsin Genetics Computer Group.

Nucleotide sequence accession number. The nucleotide sequence of the *valS* gene has been deposited in the GenBank/EMBL databases under the accession number X77239.

RESULTS

Identification of the *valS* gene. The putative *B. subtilis valS* gene encoding valyl-tRNA synthetase (ValS) had previously been identified by a homology search of a sequence upstream of the *folC* gene (encoding folyl-polyglutamate synthetase) that comprises the C-terminal 56 amino acids of a truncated open reading frame (27). Plasmid pDG1129 was constructed by P. Stragier (unpublished data) and carries a 3.15-kb fragment containing the chromosomal region immediately upstream of *folC*, including the sequence described above. We sequenced this 3.15-kb fragment and found it to contain the entire *valS* transcriptional unit. A 540-bp segment of the 5′ end of this sequence contains the *valS* leader region and is shown in Fig. 1. An open reading frame encoding a protein of 880 amino acids was identified between positions 486 and 3125 (data not shown). The deduced protein sequence shows strong similarity to the valyl-tRNA synthetases of *Bacillus stearothermophilus* (1), *Lactobacillus casei* (40), and *E. coli* (18). Sequence alignment of the four known prokaryotic synthetases shows that, as expected, the *B. subtilis* ValS is more closely related to its homologs from the gram-positive organisms *B. stearothermophilus* and *Lactobacillus casei* (89% similarity and 80% identity, and 75% similarity and 61% identity, respectively) than to the *E. coli* enzyme (67% similarity and 46% identity). The N-terminal two-thirds of the protein is well conserved between all four organisms. This part of the protein contains the catalytic core (23, 41), including the signature sequences HIGH and KMSKS of the Rossmann nucleotide binding fold in class I aaRS (6). However, the *E. coli* synthetase contains some quite extensive insertions in this part of the protein that are found in none of the other three synthetases, possibly reflecting species-related differences between gram-positive and gram-negative organisms. The *L. casei* enzyme has a 19-amino-acid N-terminal extension compared to the two *Bacillus* synthetases. The C-terminal third of the four ValS proteins is more divergent and emphasizes the close evolutionary distance between the two *Bacillus* species. The supposition initially advanced by Wetzel (42), that valyl-, isoleucyl-, leucyl-, and methionyl-tRNA synthetase are all members of a subfamily within the aaRS, is supported by aligning the sequences of these proteins from different origins (see Discussion).

We identified a potential σA-type promoter (Fig. 1) with a spacing of 17 bp and a near-consensus sequence (TTGACA and TATAAT for *B. subtilis* σ^A- and *E. coli* σ^{70}-type promoters) ~300 nucleotides upstream of the start codon. Its functionality has been confirmed, as described below. The roughly 300-bp leader contains all of the regulatory elements necessary

TABLE 1. ValS activity in total cell extracts of *valS*-overexpressing *E. coli* and *B. subtilis* cells[a]

Bacterial strain	ValS activity (pmol of charged tRNA^{Val}/μg of total protein) in:		Overexpression (fold)
	Vector (pHM3)	Vector + *valS*[b]	
E. coli KE89	1.1	6.5	6
B. subtilis SSB184[c]	11	27	2.5

[a] All data are average values from three independent experiments.
[b] pDG1129 in *E. coli*; pHMV11 in *B. subtilis*.
[c] SSB184 is the *B. subtilis* wild-type strain 1A2 (BGSC) containing the *valS-lacZ* fusion HMV8.

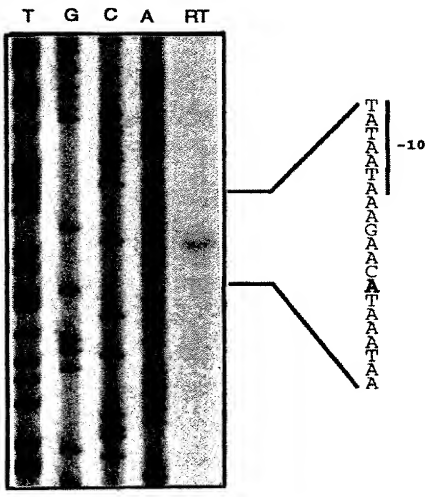


FIG. 3. Primer extension analysis of *B. subtilis valS* mRNA. Total RNA of a *B. subtilis* wild-type strain was reverse transcribed with a primer complementary to nucleotides 479 to 498 in Fig. 1. The same oligonucleotide was used for the sequencing reaction with plasmid pDG1129 as a template. RT, reverse transcription.

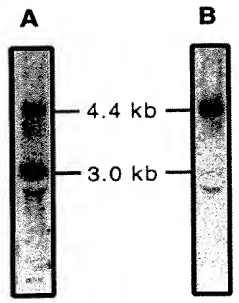


FIG. 4. Northern blot analysis of *valS* transcripts. (A) A radiolabeled 1.5-kb *Hin*dIII fragment of the *valS* structural gene (Fig. 5) was used to probe total RNA extracted from a *B. subtilis* wild-type strain. (B) The same blot was stripped of the *valS* probe and rehybridized with a *folC* PCR probe (positions 73 to 1096 of the structural gene; Fig. 5). The sizes of the two mRNA species were estimated using the BRL 0.24- to 9.5-kb RNA molecular weight marker.

to assign it to the family of genes regulated by tRNA-mediated antitermination (Fig. 1 and 2): a Rho-independent transcription terminator which is preceded by the T-box consensus sequence upstream of the structural gene and a highly structured specifier domain in the 5′ half of the leader which contains the potential GUC specifier codon. We have analyzed the importance of these elements for *valS* regulation (see below).

The *valS* gene product can charge tRNA^{Val} in vitro. In order to prove the identity of the sequenced gene, we overexpressed it in both *E. coli* and *B. subtilis* and measured an increase in tRNA^{Val} aminoacylation activity in cell extracts in vitro. For overexpression in *E. coli*, we transformed plasmid pDG1129 into the *pcnB* strain KE89, since this high-copy-number recombinant plasmid could not be stably maintained in a *pcnB*$^+$ strain. The *valS* gene was transferred to *B. subtilis* by transformation with plasmid pHMV11, constructed by inserting the 3.15-kb insert of pDG1129 into the shuttle vector pHM3. The aminoacylation activities found in the various cell extracts are given in Table 1. The 6- and 2.5-fold increases in activity in *E. coli* and *B. subtilis* cells, respectively, that were harboring the recombinant plasmids clearly show that the cloned gene encodes a functional valyl-tRNA synthetase. The difference in increase in absolute ValS activity between *E. coli* and *B. subtilis* harboring plasmids pDG1129 and pHMV11, respectively, could reflect a lower plasmid copy number in *E. coli* (*pcnB*) than in *B. subtilis* or a lower expression of the heterologous *B. subtilis* *valS* gene in *E. coli*. The 2.5-fold increase in ValS activity observed in *B. subtilis* also serves as a reference value for the autoregulation studies described below.

Mapping of *valS* transcripts. The transcription start site of *valS* was determined by primer extension analysis with an oligonucleotide complementary to nucleotides 479 to 498 in Fig. 1. Reverse transcription reactions identified a single band corresponding to a transcription start point at position 174 (Fig. 3), which is consistent with the proposed promoter.

Northern analysis of *valS* transcripts during exponential growth, using a 1.5-kb *valS* internal *Hin*dIII fragment as a probe (see Fig. 5), revealed two major transcripts of 3 and 4.4 kb (Fig. 4A). Some larger RNAs appear to be carried along in front of the 23S rRNA to give a weak additional signal. A probe specific for the *folC* gene located immediately downstream of *valS* (Fig. 5) also hybridizes to the 4.4-kb transcript but does not hybridize to the 3-kb mRNA (Fig. 4B). Thus, we believe that the 3-kb mRNA species corresponds to the *valS* mRNA and results from transcription termination at the Rho-independent terminator located in the short intergenic region between *valS* and *folC* (Fig. 5) (30, 39) and that the 4.4-kb transcript is a polycistronic mRNA comprising both the *valS* and the *folC* genes.

A GUC triplet confers the specificity of *valS* induction. The expression of the wild-type *valS* gene and that of various leader mutants was studied with the help of *lacZ* transcriptional fusions integrated in single copy at the *amy* locus of a wild-type strain or a strain auxotrophic for valine. The wild-type *valS-lacZ* fusion (HMV8) was induced almost threefold by starvation for valine. It is noteworthy that efficient valine starvation could be achieved only by adding excess leucine to the medium, despite the fact that the strain used (1A232) is not a leucine auxotroph (see Discussion). To test the relevance of the GUC triplet (Fig. 1 and 2) to *valS* induction during valine starvation, we measured β-galactosidase activity in fusions where the amino acid identity of this triplet had been changed. HMV15 has the GUC specifier codon replaced by an ACC triplet, the threonine codon which confers specificity of *thrS* induction. HMV13 contains a mutation in the T box (TGGT → TGGA) in addition to the GUC → ACC mutation to retain base pairing with the discriminator base (U) of the Thr-tRNA^{GGU} isoacceptor. HMV14 has a TAA stop codon in place of the wild-type GUC triplet. The results are summarized in Table 2. Changing the GUC (Val) to an ACC (Thr) triplet causes loss of induction by valine starvation and renders expression inducible by threonine starvation (3.4-fold). At the same time, the basal level of expression decreases more than 10-fold. Adaptation of the T-box sequence to better accommodate the interaction of the *valS* leader with the Thr-tRNA^{GGU} isoacceptor restores the basal level of expression to wild-type levels but, paradoxically, causes a near loss of induction by threonine starvation (Table 2). Replacing the GUC specifier codon with TAA (stop codon) renders the *valS* gene uninducible.

***valS* expression is autoregulated.** We previously showed that expression of *thrS* and *thrZ*, but not *pheS*, is autoregulated in a specifier-codon-dependent manner. In order to analyze whether autorepression is confined to the *thrS/thrZ* system or represents a more widespread phenomenon, we introduced the recombinant ValS overproducing plasmid, pHMV11, into a

TABLE 2. Effect of specifier codon and T-box mutations on induction of *valS-lacZ* expression

valS-lacZ fusion[a]	Specifier codon	T-box sequence	β-Galactosidase activity (U/mg)[b]				
			Complete medium	Valine starvation	Induction	Threonine starvation	Induction
HMV8 (wt)	GUC (Val)	-UGGU-	22	56	2.6×	21	0.9×
HMV15	ACC (Thr)	-UGGU-	1.6	0.6	0.4×	5.4	3.4×
HMV13	ACC (Thr)	-UGGA-	15	5	0.3×	18	1.2×
HMV14	UAA (stop)	-UGGU-	0.9	0.5	0.6×	ND	ND

[a] These fusions were measured in a strain auxotrophic for valine, strain 1A96 (see Materials and Methods). wt, wild type.
[b] ND, not done. All data represent average values from at least three independent experiments.

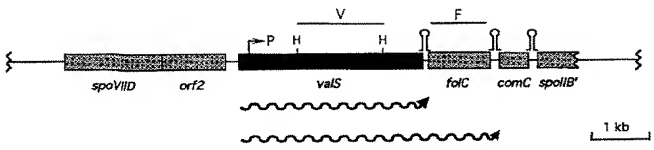


FIG. 5. Chromosomal neighborhood of the *B. subtilis valS* gene (39). The *valS* promoter and potential Rho-independent transcription terminators are indicated. Wavy arrows symbolize the mRNA species observed by Northern analysis (Fig. 4). The lines labeled V and F above the *valS* and *folC* genes indicate the sizes and positions of the fragments used as probes in Northern analysis (Fig. 4). H, *Hin*dIII.

strain carrying the wild-type *valS-lacZ* fusion (HMV8). As shown in Table 3, a 2.5-fold increase in *valS* activity was sufficient to repress the activity of the *valS-lacZ* fusion over 5-fold. Thus, expression of *valS* appears to be extremely sensitive to variations in the intracellular concentration of the synthetase. While a direct role for the synthetase in *valS* regulation cannot be ruled out at present, it appears more likely that autoregulation occurs by altering the ratio of charged to uncharged valyl-tRNA. Due to the extremely low levels of β-galactosidase expression in the stop codon mutant fusion (HMV14, Table 2) and the GUC → ACC mutant fusion (HMV15, Table 2), we could not test them for autoregulation. Therefore, we analyzed the importance of the specifier codon for autorepression in the double mutant HMV13 fusion (specifier codon and T box adapted to match the Thr-tRNA^{GGU} isoacceptor; Table 3), which has a higher basal level of expression (Table 2). The double mutation led to a loss of autoregulation (1.7-fold repression), underlining the importance of these two sites of tRNA-mRNA interaction for this type of regulation. Although no repression was observed with a 10-fold overproduction of ThrS (Table 3), this is perhaps not surprising given that the HMV13 fusion is also not inducible by threonine starvation (see above).

DISCUSSION

The three valyl-tRNA synthetases from the gram-positive organisms *B. stearothermophilus* (1), *Lactobacillus casei* (40), and *B. subtilis* are very similar and more compact than their *E. coli* counterpart (18), which contains some extensive insertions in the amino-terminal two-thirds of the protein. Comparison of the *B. subtilis* ValS sequence with other branched-chain aaRS proteins in bacteria revealed surprisingly strong similarities between *B. subtilis* ValS and the following synthetases (expressed in percentage per unit length, similarity and identity): *B. subtilis* MetS, 49.2 and 26.2%; *E. coli* MetS, 45.6 and 21.4%; *E. coli* IleS 49.5 and 26.3%; *B. subtilis* LeuS, 50.9 and 25.2%; and *E. coli* LeuS, 54.3 and 29.8%. It is interesting that similarities between heterologous synthetases are not necessarily higher when they originate from the same organism (*B. subtilis* ValS is 50.9% similar to *B. subtilis* LeuS but 54.3% similar to *E. coli* LeuS), implying that the common ancestor of the branched-chain aaRS probably existed before the separation of bacteria in a gram-positive and gram-negative kingdom.

Northern blot analysis revealed the presence of two transcripts (3 and 4.4 kb) containing the *valS* gene. The 4.4-kb transcript also hybridized to a *folC*-specific probe, indicating that both genes are likely to be cotranscribed on a polycistronic mRNA originating at the *valS* promoter. This is also the predicted length of a transcript extending from the *valS* promoter to the transcription terminator immediately downstream of *folC*. The presence of roughly equal amounts of the two mRNAs indicates that the *valS* terminator is only about 50% efficient. Overexposure of the Northern blot revealed only very low quantities of transcripts extending beyond the *folC* transcription terminator. This is consistent with the finding that expression of *comC*, the gene lying downstream of *folC* (Fig. 5), is induced only during late competence (30).

We attempted to inactivate the *valS* gene on the chromosome and found this not be lethal. Disruption of *valS* in the survivors was confirmed by Southern blotting (data not shown) and suggests that a second functional gene with valine-tRNA synthetase activity exists in *B. subtilis*, as is the case for the threonyl- and tyrosyl-tRNA synthetases.

Sequence and two-dimensional structure analyses of the *valS* leader suggested that this gene is a member of the family of genes in gram-positive organisms that comprises aaRS and amino acid biosynthetic genes regulated by tRNA-mediated antitermination (14). Expression of *valS* was induced by starvation for valine, but this derepression could be observed only when the cells were grown in the presence of excess leucine, despite the fact that the *trpC2 ilvD4* mutant strain used in this study is auxotrophic for tryptophan, isoleucine, and valine but not for leucine. A rationale for this observation may be found in the way the *ilv-leu* biosynthetic operon is regulated. Expression of the *ilv-leu* operon is also likely to be regulated by the level of charged/uncharged Leu-tRNA via tRNA-mediated antitermination (28) and responds to variations in leucine concentration (10). Since the uncharacterized *ilvD4* mutation used here shows a slightly leaky phenotype, we believe that excess leucine further shuts down *ilv-leu* expression, thereby creating conditions whereby valine starvation can occur more efficiently.

The specificity of *valS* induction depends on the identity of the strategically placed GUC (Val) triplet (specifier codon) in the extensive 5′-terminal secondary structure (Fig. 2). Changing the GUC triplet to ACC (Thr) switched the specificity of induction from valine to threonine starvation. However, the GUC → ACC transition leads to a more than 10-fold drop in the basal level of expression, very close to the activity of an uninducible fusion in which the GUC specifier was mutated to a UAA stop codon. Clearly, the Thr-tRNA^{GGU} isoacceptor interacts much less efficiently with the *valS* leader containing the ACC codon than Val-tRNA interacts with the wild-type leader. In order to improve this interaction, we altered the T-box sequence to match the discriminator base of the Thr-

TABLE 3. Effect of *valS* and *thrS* overexpression on wild-type and mutant *valS-lacZ* fusions

valS-lacZ fusion	Specifier codon	T-box sequence	Multicopy plasmid	Insert	β-Galactosidase sp act (U/mg)[a]	Repression factor (fold)	Overexpression of synthetase (fold)
HMV8 (wt)[b]	GUC	-UGGU-	pHM3	Control	39		
			pHMV11	*valS*	7.5	5.2	2.5
HMV13	ACC	-UGGA-	pHM3	Control	44		
			pHMV11	*valS*	26	1.7	2.5

[a] All β-galactosidase activities are average values from at least three independent experiments.
[b] wt, wild type.

$tRNA^{GGU}$ isoacceptor. Indeed, basal expression rose about 10-fold and approached wild-type levels but at the same time became almost uninducible by threonine starvation (Table 2). This phenomenon is difficult to explain. The GUC → ACC transition (HMV15) was sufficient to render the mutated *valS-lacZ* fusion inducible by starvation for threonine, indicating that the Thr-$tRNA^{GGU}$ isoacceptor can, albeit not very efficiently, interact with the *valS* leader and recognize the ACC specifier codon. Permitting the T-box sequence to base pair with the discriminator base of the Thr-$tRNA^{GGU}$ (HMV13) seemed to improve the tRNA-mRNA interaction, as reflected by a 10-fold increase in basal expression. However, if this interaction is indeed so efficient, one would expect this mutant fusion to be highly inducible by threonine starvation, which is clearly not the case. We previously encountered a similar phenomenon when introducing analogous mutations in the *thrS* leader to complement the discriminator base of Phe-tRNA (35).

It seems logical that genes whose expression responds to the ratio of charged to uncharged cognate tRNA would also be affected by the intracellular concentration of their product responsible for charging these tRNAs, but of the two systems studied to date (*thrS/thrZ* [8] and *pheS* [35]), only the *thrS* and *thrZ* genes were autoregulated. To see whether this type of regulation represents a more common phenomenon was one of the reasons we tested whether *valS* expression is autoregulated. As shown in this study, a 2.5-fold overexpression of *valS* from a multicopy plasmid led to a more than 5-fold repression of a wild-type *valS-lacZ* fusion. Expression of the *valS* gene is thus more tightly controlled by the intracellular concentration of its product than is the case for *thrS*, where 10-fold overproduction leads to 10-fold repression (8). This could be explained if basal *valS* expression were to depend much more heavily on antitermination mediated by uncharged tRNA than is the case for *thrS*. In minimal medium, expression of a *valS-lacZ* fusion, in which the GUC specifier codon has been changed to a TAA stop codon, drops 20-fold compared to expression of the wild-type fusion (Table 2), while an equivalent change in the *thrS* system leads to only a 2-fold drop in expression (35). While it seems likely that autorepression occurs by altering the charged/uncharged tRNA ratio in the cell, an additional direct interaction of the synthetase with the mRNA cannot yet be excluded.

ACKNOWLEDGMENTS

We thank C. Condon for helpful discussions and critical reading of the manuscript.

This work was supported by grants from the CNRS (UPR9073) and the European Union (ERBCHRX-CT94-0529 to M.G.-M.) and by fellowships from the CNRS and the Société de Secours des Amis des Sciences (to D.L.).

REFERENCES

1. **Brand, N. J., and A. R. Fersht.** 1986. Molecular cloning of the gene encoding the valyl-tRNA synthetase from *Bacillus stearothermophilus*. Gene **44**:139–142.
2. **Bruand, C., A. Sorokin, P. Serror, and S. D. Ehrlich.** 1995. Nucleotide sequence of the *Bacillus subtilis dnaD* gene. Microbiology **141**:321–322.
3. **Chambers, S. P., S. E. Prior, D. A. Barstow, and N. P. Minton.** 1988. The pMTL *nic*⁻ cloning vectors. I. Improved pUC polylinker regions to facilitate the use of sonicated DNA for nucleotide sequencing. Gene **68**:139–149.
4. **Condon, C., M. Grunberg-Manago, and H. Putzer.** 1996. Aminoacyl-tRNA synthetase gene regulation in *Bacillus subtilis*. Biochimie **78**:381–389.
5. **Condon, C., H. Putzer, and M. Grunberg-Manago.** 1996. Processing of the leader mRNA plays a major role in the induction of *thrS* expression following threonine starvation in *B. subtilis*. Proc. Natl. Acad. Sci. USA **93**:6992–6997.
6. **Eriani, G., M. Delarue, O. Poch, J. Gangloff, and D. Moras.** 1990. Partition of transfer RNA synthetases into 2 classes based on mutually exclusive sets of sequence motifs. Nature **347**:203–206.
7. **Gagnon, Y., R. Breton, H. Putzer, M. Pelchat, M. Grunberg-Manago, and J. Lapointe.** 1994. Clustering and co-transcription of the *Bacillus subtilis* genes encoding the aminoacyl-tRNA synthetases specific for glutamate and for cysteine and the first enzyme for cysteine biosynthesis. J. Biol. Chem. **269**:7473–7482.
8. **Gendron, N., H. Putzer, and M. Grunberg-Manago.** 1994. Expression of both *Bacillus subtilis* threonyl-tRNA synthetase genes is autogenously regulated. J. Bacteriol. **176**:486–494.
9. **Glaser, P., A. Danchin, F. Kunst, M. Débarbouillé, A. Vertès, and R. Dedonder.** 1990. A gene encoding a tyrosine-tRNA synthetase is located near *sacS* in Bacillus subtilis. J. DNA Mapping Sequencing **1**:251–261.
10. **Grandoni, J. A., S. B. Fullmer, V. Brizzio, S. A. Zahler, and J. M. Calvo.** 1993. Regions of the *Bacillus subtilis ilv-leu* operon involved in regulation by leucine. J. Bacteriol. **175**:7581–7593.
11. **Grandoni, J. A., S. A. Zahler, and J. M. Calvo.** 1992. Transcriptional regulation of the *ilv-leu* operon of *Bacillus subtilis*. J. Bacteriol. **174**:3212–3219.
12. **Grunberg-Manago, M.** 1996. Regulation of the expression of aminoacyl-tRNA synthetases and translation factors, p. 1432–1457. *In* F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2nd ed., vol. 1. ASM Press, Washington, D.C.
13. **Grundy, F. J., and T. M. Henkin.** 1994. Conservation of a transcription antitermination mechanism in aminoacyl-tRNA synthetase and amino acid biosynthesis genes in Gram-positive bacteria. J. Mol. Biol. **235**:798–804.
14. **Grundy, F. J., and T. M. Henkin.** 1993. tRNA as a positive regulator of transcription antitermination in *B. subtilis*. Cell **74**:475–482.
15. **Grundy, F. J., S. M. Rollins, and T. M. Henkin.** 1994. Interaction between the acceptor end of tRNA and the T box stimulates antitermination in the *Bacillus subtilis tyrS* gene: a new role for the discriminator base. J. Bacteriol. **176**:4518–4526.
16. **Guerout-Fleury, A. M., K. Shazand, N. Frandsen, and P. Stragier.** 1995. Antibiotic-resistance cassettes for *Bacillus subtilis*. Gene **167**:335–336.
17. **Harwood, C. R., and S. M. Cutting (ed.).** 1990. Molecular biological methods for bacillus. John Wiley & Sons Ltd., Chichester, England.
18. **Heck, J. D., and W. G. Hatfield.** 1988. Valyl-tRNA synthetase gene of *E. coli* K12. J. Biol. Chem. **263**:857–867.
19. **Henkin, T. M.** 1994. tRNA-directed transcription antitermination. Mol. Microbiol. **13**:381–387.
20. **Henkin, T. M., B. L. Glass, and F. J. Grundy.** 1992. Analysis of the *Bacillus subtilis tyrS* gene: conservation of a regulatory sequence in multiple tRNA synthetase genes. J. Bacteriol. **174**:1299–1306.
21. **Hirshfield, I. N., P. L. Bloch, R. A. VanBogelen, and F. C. Neidhardt.** 1981. Multiple forms of lysyl-transfer ribonucleic acid synthetase in *Escherichia coli*. J. Bacteriol. **146**:345–351.
22. **Hirshfield, I. N., R. Tenreiro, R. A. VanBogelen, and F. C. Neidhardt.** 1984. *Escherichia coli* K-12 lysyl-tRNA synthetase mutant with a novel reversion pattern. J. Bacteriol. **158**:615–620.
23. **Hountondji, C., P. Dessen, and S. Blanquet.** 1986. Sequence similarities among the family of aminoacyl-tRNA synthetases. Biochimie **68**:1071–1078.
24. **Kunkel, T. A., J. D. Roberts, and R. A. Zakour.** 1987. Rapid and efficient site-specific mutagenesis without phenotypic selection, p. 367–382. *In* R. Wu and L. Grossman (ed.), Recombinant DNA, part E, vol. 154. Academic Press, New-York.
25. **Kunst, F., M. Debarbouille, T. Msadek, M. Young, C. Mauel, D. Karamata, A. Klier, G. Rapoport, and R. Dedonder.** 1988. Deduced polypeptides encoded by the *Bacillus subtilis sacU* locus share homology with two-component sensor-regulator systems. J. Bacteriol. **170**:5093–5101.
26. **Lévêque, F., P. Plateau, P. Dessen, and S. Blanquet.** 1990. Homology of *lysS* and *lysU*, the two *E. coli* genes encoding distinct lysyl-tRNA synthetase species. Nucleic Acids Res. **18**:305–312.
27. **Margolis, P. S., A. Driks, and R. Losick.** 1993. Sporulation gene *spoIIB* from *Bacillus subtilis*. J. Bacteriol. **175**:528–540.
28. **Marta, P. T., R. D. Ladner, and J. A. Grandoni.** 1996. A CUC triplet confers leucine-dependent regulation of the *Bacillus subtilis ilv-leu* operon. J. Bacteriol. **178**:2150–2153.
29. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
30. **Mohan, S., J. Aghion, N. Guillen, and D. Dubnau.** 1989. Molecular cloning and characterisation of *comC*, a late competence gene of *Bacillus subtilis*. J. Bacteriol. **171**:6043–6051.
31. **Ogasawara, N., S. Nakai, and H. Yoshikawa.** 1994. Systematic sequencing of the 180 kilobase region of the *Bacillus subtilis* chromosome containing the replication origin. DNA Res. **1**:1–14.
32. **Putzer, H., A. A. Brackhage, and M. Grunberg-Manago.** 1990. Independent genes for two threonyl-tRNA synthetases in *Bacillus subtilis*. J. Bacteriol. **172**:4593–4602.
33. **Putzer, H., N. Gendron, and M. Grunberg-Manago.** 1992. Co-ordinate expression of the two threonyl-tRNA synthetase genes in *Bacillus subtilis*: control by transcriptional antitermination involving a conserved regulatory sequence. EMBO J. **11**:3117–3127.
34. **Putzer, H., M. Grunberg-Manago, and M. Springer.** 1995. Bacterial amino-

acyl-tRNA synthetases: genes and regulation of expression, p. 293–333. *In* D. Söll and U. L. RajBhandary (ed.), tRNA: structure, biosynthesis, and function. American Society for Microbiology, Washington, D.C.

35. **Putzer, H., S. Laalami, A. A. Brakhage, C. Condon, and M. Grunberg-Manago.** 1995. Aminoacyl-tRNA synthetase gene regulation in *Bacillus subtilis*: induction, repression and growth-rate regulation. Mol. Microbiol. **16:** 709–718.
36. **Renault, P., J. J. Godon, N. Goupil, C. Delorme, G. Corthier, and S. D. Ehrlich.** 1995. Metabolic operons in *Lactococci*. Dev. Biol. Stand. **85:**431–441.
37. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
38. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
39. **Schmidt, R., A. L. Decatur, P. H. Rather, P. C. Moran, and R. Losick.** 1994. *Bacillus subtilis* Lon protease prevents inappropriate transcription of genes under the control of the sporulation transcription factor σG. J. Bacteriol. **176:**6528–6537.
40. **Taylor, B. V., J. Toy, T. L. Sit, and A. L. Bognar.** 1993. Cloning and sequence determination of the *valS* gene, encoding valyl-tRNA synthetase in *Lactobacillus casei*. J. Bacteriol. **175:**2475–2478.
41. **Webster, T., H. Tsai, M. Kula, G. A. Mackie, and P. Schimmel.** 1984. Specific sequence homology and three-dimensional structure of an aminoacyl transfer RNA synthetase. Science **226:**1315–1317.
42. **Wetzel, R.** 1978. Aminoacyl-tRNA synthetase families and their significance to the origin of the genetic code. Origins Life **9:**39–50.
43. **Wilcox, M.** 1969. γ-Phosphoryl ester of Glu-tRNA^{Gln} as an intermediate in *Bacillus subtilis* glutaminyl-tRNA synthesis. Cold Spring Harbor Symp. Quant. Biol. **34:**521–528.